INEQUALITY AND GROWTH: PATTERNS AND POLICY

Volume I: Concepts and Analysis

Edited by
Kaushik Basu
and Joseph E. Stiglitz

IEA Conference Volume No. 156-I

International Economic Association

INEQUALITY AND GROWTH: PATTERNS AND POLICY
Volume 1: Concepts and Analysis
Edited by Kaushik Basu and Joseph E. Stiglitz

INEQUALITY AND GROWTH: PATTERNS AND POLICY
Volume 2: Regions and Regularities
Edited by Kaushik Basu and Joseph E. Stiglitz

TOWARDS A GENERAL THEORY OF DEEP DOWNTURNS
Written by Joseph E. Stiglitz

CONTEMPORARY ISSUES IN MACROECONOMICS
Lessons from the Crisis and Beyond
Edited by Joseph E. Stiglitz and Martin Guzman

CONTEMPORARY ISSUES IN MICROECONOMICS
Edited by Joseph E. Stiglitz and Martin Guzman

CONTEMPORARY ISSUES IN DEVELOPMENT ECONOMICS
Edited by Timothy Besley

THE MIDDLE EAST ECONOMIES IN TIMES OF TRANSITION
Edited by Ishac Diwan and Ahmed Galal

TAMING CAPITAL FLOWS
Capital Account Management in an Era of Globalization
Edited by Joseph E. Stiglitz and Refet Gurkaynak

LIFE AFTER DEBT
The Origins and Resolutions of Debt Crisis
Edited by Joseph E. Stiglitz and Daniel Heymann

INCOME CONTINGENT LOANS
Theory, Practice and Prospects
Edited by Joseph E. Stiglitz, Bruce Chapman and Timothy Higgins

THE INDUSTRIAL POLICY REVOLUTION I
The Role of Government Beyond Ideology
Edited by Joseph E. Stiglitz and Justin Lin Yifu

THE INDUSTRIAL POLICY REVOLUTION II
Africa in the 21st Century
Edited by Joseph E. Stiglitz, Justin Lin Yifu and Ebrahim Patel

THE CHINESE ECONOMY
A New Transition
Edited by Masahiko Aoki and Jinglian Wu

INSTITUTIONS AND COMPARATIVE ECONOMIC DEVELOPMENT
Edited by Franklin Allen, Masahiko Aoki, Nobuhiro Kiyotaki, Roger Gordon, Joseph E. Stiglitz and Jean-Paul Fitoussi

COMPLEXITY AND INSTITUTIONS: MARKETS, NORMS AND CORPORATIONS
Edited by Masahiko Aoki, Kenneth Binmore, Simon Deakin and Herbert Gintis

CORPORATE SOCIAL RESPONSIBILITY AND CORPORATE GOVERNANCE
The Contribution of Economic Theory and Related Disciplines
Edited by Lorenzo Sacconi, Margaret Blair, R. Edward Freeman and Alessandro Vercelli

IS ECONOMIC GROWTH SUSTAINABLE?
Edited by Geoffrey Heal

KEYNES'S GENERAL THEORY AFTER SEVENTY YEARS
Edited by Robert Diman, Robert Mundell and Alessandro Vercelli

CORRUPTION, DEVELOPMENT AND INSTITUTIONAL DESIGN
Edited by János Kornai, László Mátyás and Gérard Roland

MARKET AND SOCIALISM
In the Light of the Experience of China and Vietnam
Edited by János Kornai and Yingyi Quian

INSTITUTIONAL CHANGE AND ECONOMIC BEHAVIOUR
Edited by János Kornai, László Mátyás and Gérard Roland

INTERGENERATIONAL EQUITY AND SUSTAINABILITY
Edited by John E. Roemer and Kotaro Suzumura

PSYCHOLOCY, RATIONALITY AND ECONOMIC BEHAVIOUR
Challenging Standard Assumptions
Edited by Bina Agarwal and Alessandro Vercelli

MULTINATIONALS AND FOREIGN INVESTMENT IN ECONOMC DEVELOPMENT
Edited by Edward M. Graham

POST-CONFLICT ECONOMIES IN AFRICA
Edited by Paul Collier and Augustin Kwasi Fosu

STRUCTURAL REFORM AND MACROECONOMIC POLICY
Edited by Robert M. Solow

THE PAST, PRESENT AND FUTURE OF THE EUROPEAN UNION
Edited by Alan V. Deardorff

LATIN AMERICAN ECONOMIC CRISES
Trade and Labour
Edited by Enrique Bour, Daniel Heymann and Fernando Navajas

ADVANCES IN MACROECONOMIC THEORY
Edited by Jacques H, Drèze

EXPLAINING GROWTH
A Global Research Project
Edited by Gary McMahon and Lyn Squire

TRADE, INVESTMENT, MIGRATION AND LABOUR MARKET ADJUSTMENT
Edited by David Greenaway, Richard Upward and Katherine Wakelin

INEQUALITY AROUND THE WORLD
Edited by Richard B. Freeman

MONETARY THEORY AND POLICY EXPERIENCE
Edited by Axel Leijonhufvud

MONETARY THEORY AS A BASIS FOR MONETARY POLICY
Edited by Axel Leijonhufvud

ECONOMIC DEVELOPMENT IN SUBSAHARAN AFRICA
Proceedings of the Eleventh World Congress of the International Economic Association, Tunis
Edited by Ibrahim Elbadawi and Beno Ndula

International Economics Association
Series Standing Order ISBN 978–0–3337–1242–9 (Hardback)
978–0–3338–0330–1 (Paperback)
(*outside North America only*)

You can receive future titles in this series as they are published by placing a standing order. Please contact your bookseller or, in case of difficulty, write to us at the address below with your name and address, the title of the series and the ISBN quoted above.

Customer Services Department, Macmillan Distribution Ltd, Houndmills, Basingstoke, Hampshire RG21 6XS, England

---

Inequality and Growth: Patterns and Policy

Volume I: Concepts and Analysis

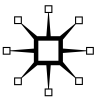
Edited by

Kaushik Basu
Chief Economist and Senior Vice President, World Bank, Washington DC, USA

and

Joseph E. Stiglitz
University Professor, Columbia University, New York, USA

palgrave
macmillan



INEQUALITY AND GROWTH: PATTERNS AND POLICY: VOLUME I: CONCEPTS AND ANALYSIS

First published 2016 by
PALGRAVE MACMILLAN

Palgrave Macmillan in the UK is an imprint of Macmillan Publishers Limited, registered in England, company number 785998, of Houndmills, Basingstoke, Hampshire RG21 6XS.

Palgrave Macmillan in the US is a division of Nature America, Inc., One New York Plaza, Suite 4500 New York, NY 10004-1562.

Palgrave Macmillan is the global academic imprint of the above companies and has companies and representatives throughout the world.

ISBN 978-1-137-55453-6 ISBN 978-1-137-55454-3 (eBook)
DOI: 10.1057/9781137554543

Distribution in the UK, Europe and the rest of the world is by Palgrave Macmillan®, a division of Macmillan Publishers Limited, registered in England, company number 785998, of Houndmills, Basingstoke, Hampshire RG21 6XS.

Library of Congress Cataloging-in-Publication Data
Names: Basu, Kaushik, | Stiglitz, Joseph E., editor. editor.
Title: Inequality and growth: Patterns and Policy, Volume I: concepts and analysis / edited by Joseph E. Stiglitz, Kaushik Basu.
Description: New York: Palgrave Macmillan, 2016. | Series: International economic association | Includes index.
Identifiers: LCCN 2015039479| ISBN 9781137554529 (hardback) | ISBN 781137554536 (paperback) | ISBN 9781137554550 (ePub) | ISBN 9781137554543 (PDF)
Subjects: LCSH: Income distribution. | Economic development. | BISAC: BUSINESS & ECONOMICS / Economic Conditions. | BUSINESS & ECONOMICS / Economics / Macroeconomics. | BUSINESS & ECONOMICS / Development / Economic Development.
Classification: LCC HC79.I5 I4934 2016 | DDC 339.2—dc23
LC record available at http://lccn.loc.gov/2015039479

A catalog record for this book is available from the Library of Congress

A catalogue record for the book is available from the British Library

Typeset by MPS Limited, Chennai, India.

Contents

List of Figures

List of Tables

Acknowledgments

The editors are most grateful to the illustrious participants of the roundtable for their contribution to this two-volume monograph. They are indebted to the staff at the IEA Secretariat for their help in the organization of the roundtable as part of the IEA World Congress meetings. They would also like to acknowledge the role of Palgrave Macmillan in shepherding the publication of the volumes and express a special thanks to Timothy Besley, President of the IEA and the program chair of the IEA 17th World Congress, for his support and advice on the organization of the roundtable. Finally, the project would not have been possible but for the support of Bintao Wang, Eamon Kircher, Debarati Ghosh, and, especially, Vivian Hon, who not only helped organize the roundtable but read through and helped with the editing work for the entire two-volume book.

Kaushik Basu
Joseph E. Stiglitz

Notes on Contributors

Editors

Kaushik Basu is the Chief Economist and Senior Vice President of the World Bank. Prior to joining the World Bank, he served as Chief Economic Adviser to the Government of India and is currently on leave from Cornell where he is Professor of Economics and the C. Marks Professor of International Studies. He holds a PhD in Economics from the London School of Economics, and is a recipient of one of India's highest civilian honors, the Padma Bhushan. His contributions span development economics, welfare economics, industrial organization and game theory. He has taught at the Delhi School of Economics, Harvard, Princeton and MIT. His most recent book is *An Economist in the Real World: The Art of Policymaking in India* (2015).

Joseph E. Stiglitz is University Professor at Columbia University, the winner of the 2001 Nobel Memorial Prize in Economics, and a lead author of the 1995 IPCC report, which shared the 2007 Nobel Peace Prize. He was chairman of the US Council of Economic Advisors under President Clinton and chief economist and senior vice president of the World Bank for 1997–2000. Stiglitz received the John Bates Clark Medal, awarded annually to the American economist under 40 who has made the most significant contribution to the subject. He was a Fulbright Scholar at Cambridge University, held the Drummond Professorship at All Souls College Oxford, and has also taught at MIT, Yale, Stanford, and Princeton.

Chapter Authors

Sudhir Anand is Professor of Economics at the University of Oxford, and Adjunct Professor of Global Health at the Harvard School of Public Health. His research interests include: economics of inequality and poverty, undernutrition, human development, and population health. He is currently researching theoretical, empirical and conceptual issues in health inequality. The economic approach to measuring inequality is adapted and developed for demographic concepts such as the life table and the survival curve which are used to estimate life expectancy. He has done empirical work on length-of-life distribution measures to generate a set of comparable estimates across countries and over time (including for France for each year during 1806–1995). Conceptual research is on health equity and counterfactuals in health evaluation. Other research is on the extent to which human resources for health matter in explaining health outcomes, and on what we know about global income

inequality and the global distribution of income. He is teaching a joint course at Harvard University with Amartya Sen in Fall 2015 on "Health, Inequality and Development."

Michele Battisti is Associate Professor of Economics at the University of Palermo and professor at Luiss University in Rome. His research interests include income convergence and inequality, the economics of organized crime and the impact on labour market institutions on productivity.

François Bourguignon is Emeritus Professor of Economics at the Paris School of Economics. He is a specialist in the economics of development, public policy, income distribution and inequality and has authored many academic papers and books. Earlier in his career, he held the position of Senior Vice President for Development Economics and Chief Economist at the World Bank (2003–07). He was then appointed as the director of the newly created Paris School of Economics, an appointment he held until 2013. Along with his present duties as an academic in a leading institution in economic teaching and research, François Bourguignon is active in the international development community by lecturing and writing reports for leading international agencies.

Tony Castleman is a development economist whose research focuses on nutrition and food security and on the roles that respect, dehumanization, and human recognition play in public health, poverty and conflict. He currently serves as Country Representative, India at Catholic Relief Services. Previously he was Associate Director of the Institute for International Economic Policy, and Associate Research Professor, at George Washington University. Prior to joining GWU, Castleman worked for the USAID-funded Food and Nutrition Technical Assistance Projects at the Academy for Educational Development, most recently as Deputy Director – Field Support. At FANTA/FANTA-2, he provided technical assistance to USAID, governments, and implementing organizations to strengthen nutrition and food security policies and programs. He led a team of specialists working in 15 countries and globally in the areas of nutrition and HIV, management of acute malnutrition, and maternal and child health and nutrition. Prior to FANTA, he was Director of an NGO in India, where he designed and managed a range of health, nutrition, education and poverty alleviation projects. He holds a Ph.D. in Economics from George Washington University and a B.A. in Public Policy from Stanford University.

Hai-Anh H. Dang is an economist in the Poverty and Inequality Unit, Development Research Group, World Bank. His main research is on development, poverty, education, and labor. His recent work focuses on methodology to construct synthetic (pseudo) panel data from cross-sections that can be employed to study welfare dynamics. He has published on these topics in

different journals as well as a book on private tutoring in Vietnam. He holds a PhD in Applied Economics from the University of Minnesota.

Bhaskar Dutta is Professor University of Warwick, UK and Ashoka University, India. His research interests include formation of groups and networks; mechanism design; cooperative game theory; political economy, and social choice theory. He is a Fellow of the Econometric Society as well as of the Society for the Advancement of Economic Theory. He is the President of the Society for Social Choice and Welfare (2014–16), and a member of the Council of Game Theory Society.

James E. Foster is Professor of Economics and International Affairs at The George Washington University. He received his PhD in economics from Cornell University, holds a Doctorate Honoris Causa, from Universidad Autónoma del Estado de Hidalgo (Mexico). His research broadly lies within welfare economics, using economic tools to evaluate and enhance the wellbeing of people. His highly cited 1984 *Econometrica* paper with Joel Greer and Erik Thorbecke introduced the FGT poverty indices, which have been used in thousands of studies and were the basis for targeting the Progresa/Oportunidades program in México. Other research includes work on multidimensional poverty with Sabina Alkire; on literacy with Kaushik Basu; on economic inequality with Amartya Sen; and on inclusive growth with Miguel Szekely.

Paola Giuliano is an Associate Professor of Economics at UCLA Anderson School of Management. Her main areas of research are culture and economics and political economy. Recent projects focus on questions related to the role of culture and history in the determination of economic outcomes, the importance of the family in the transmission of economic values and the effect of recessions on preferences for redistribution and political behavior. She is a faculty research fellow of the NBER and a research affiliate of the CEPR (London) and IZA (Institute for the Study of Labor). She holds a BA from Bocconi University (Milan) and a PhD in Economics from the University of California at Berkeley. She is a co-editor of the *Journal of the European Economic Association*.

Karla Hoff is Co-Director of the *World Development Report* 2015 on Mind, Society, and Behavior and a Senior Research Economist at the World Bank. Most of her work has been on social, cultural, and institutional obstacles to development and on policies to overcome them. Her papers in the *American Economic Review* explain how inefficient neighborhoods emerge with segregation between renters and homeowners; how cues to a stigmatized identity impair performance, which can make false stereotypes appear to be true ("equilibrium fictions"); and how the asset stripping unleashed by the "Big Bang" privatization in Russia produced a coordination failure that reduced the demand for a broadly beneficial rule of law. She won an award for one of the

top 50 economics papers for "Exiting a Lawless State," co-authored with Joseph Stiglitz. She has taught at the University of Maryland, Princeton University, and the University of Pennsylvania.

Vivian Hon is Adviser to the Chief Economist and Senior Vice President of the World Bank. She joined the World Bank in 1998 as an economist in the Economics Policy Department of the Poverty Reduction and Economics Network. Prior to that, she was pursuing her PhD. in Economics at the University of North Carolina and working as a consultant in the Human Development Department of the World Bank. She has worked in development economics in the areas of fiscal decentralization, labor economics, economic geography and health financing. She was part of the core team operationalizing the *World Development Report 2009* on "Reshaping Economic Geography." She holds an MSc. In Economics from the University of North Carolina at Chapel Hill.

John Ifcher is an Assistant Professor of Economics at Santa Clara University's Leavey School of Business. His recent research focuses on subjective wellbeing, time preference, and social welfare programs. One recent paper investigates the impact of positive affect on time preference; another investigates the effect of the Earned Income Tax Credit on mental health and happiness. He holds a PhD in Economics from the University of California at Berkeley. He also received a MPA from the School of International and Public Affairs at Columbia University.

Ravi Kanbur is T. H. Lee Professor of World Affairs, International Professor of Applied Economics and Management, and Professor of Economics at Cornell University. He holds an appointment tenured both in the Charles H. Dyson School of Applied Economics and Management in the College of Agriculture and Life Sciences, and in the Department of Economics in the College of Arts and Sciences. His main areas of interest are public economics and development economics. His work spans conceptual, empirical, and policy analysis. He is particularly interested in bridging the worlds of rigorous analysis and practical policymaking. His vita lists over 230 publications, covering topics such as risk taking, inequality, poverty, structural adjustment, labor, urbanization and agriculture. The honors he has received include the Quality of Research Discovery Award of the American Agricultural Economics Association and an Honorary Professorship at the University of Warwick. He is the current President of the Society for the Study of Economic Inequality, and a member of the OECD High Level Expert Group on the Measurement of Economic Performance and Social Progress.

Peter Lanjouw is Professor of Development Economics at VU University, Amsterdam. Prior to this appointment he was Research Manager of the Poverty Team in the Development Economics Research Group of the World Bank. He joined the World Bank in 1992, after completing his doctoral

studies in economics from the London School of Economics. His research focuses on the analysis of poverty and inequality as well as on rural development, notably the study of a village economy in rural India and the broader analysis of rural non-farm diversification. He has taught at UC Berkeley, the University of Namur, and the Foundation for the Advanced Study of International Development in Tokyo, and he is also an Honorary Fellow of the Amsterdam Institute of International Development, Amsterdam. His articles have appeared in *Econometrica, The Economic Journal* and the *Journal of Public Economics* as well as numerous field journals such as the *Journal of Development Economics, The World Bank Economic Review, World Development, Agricultural Economics* and the *Journal of Economic Inequality*. He is currently editor of the *World Bank Research Observer.*

Murray Leibbrandt is the Pro Vice-Chancellor, Poverty and Inequality, at the University of Cape Town, a Professor in the School of Economics and the Director of the Southern Africa Labour and Development Research Unit. He holds the DSD/NRF National Research Chair of Poverty and Inequality Research and is an IZA Research Fellow. His research analyses South African poverty, inequality and labor market dynamics using survey data and, in particular, panel data. He is one of the Principal Investigators on the National Income Dynamics Study.

Martin Ravallion holds the Edmond D. Villani Chair of Economics at Georgetown University. Prior to joining Georgetown in December 2012 he was Director of the World Bank's research department, the Development Research Group. He joined the Bank in 1988 and has worked in almost all sectors and all regions over the following 24 years. Prior to joining the Bank, he was on the faculty of the Australian National University. He holds a PhD in Economics from the London School of Economics, and has taught economics at LSE, Oxford University, the Australian National University and Princeton University. His main research interests over the last 25 years have concerned poverty and policies for fighting it. He currently serves on the Editorial Boards of ten economics journals, is a Senior Fellow of the Bureau for Research in Economic Analysis of Development, a Founding Council Member and President of the Society for the Study of Economic Inequality, a Research Associate of the National Bureau of Economic Research and a Non-Resident Fellow of the Center for Global Development. Amongst various prizes and awards, in 2012 he was awarded the John Kenneth Galbraith Prize from the American Agricultural and Applied Economics Association.

Stephen C. Smith is Professor of Economics and International Affairs at George Washington University. He received his PhD in Economics from Cornell University and has been a Fulbright Research Scholar, a Fulbright Senior

Specialist, a Visiting Fellow at the Brookings Institution, and a Jean Monnet Research Fellow; he is currently Nonresident Senior Fellow at the Brookings Institution, an IZA Research Fellow, and a member of the Advisory Council of BRAC USA. Professor Smith has worked in several regions of the developing world and has been a consultant for the World Bank, the International Labour Organization, the United Nations Development Programme, and the World Institute for Development Economics Research. He is Principal Investigator for a research project studying training, technology, and credit complementarities in smallholder agriculture in Senegal and Uganda, funded by BASIS/USAID. He has published widely in the fields of economic development, poverty, and the economics of organization and participation. Smith is co-author (with Michael Todaro) of the widely used textbook, *Economic Development*, now in its 12th edition (2014).

Aristomene Varoudakis is Professor of Economics at the University of Strasbourg, France, where he teaches at the Institut d'Etudes Politiques and the Department of Economics. Previously he served as a Lead Economist and Country Manager at the World Bank's Europe and Central Asia and Middle East and North Africa Departments and as an adviser at the World Bank's Independent Evaluation Group. Prior to joining the World Bank, he served as Senior Economist at the OECD Economics Department and the OECD Development Center. His research interests include macroeconomics and growth, development economics, public finance, international finance and exchange rates. His work has been published in academic journals and collective volumes. He is the author or co-author of four books.

Adam Wagstaff is Research Manager of the Human Development and Public Services team in the World Bank's Development Research Group. He has extensive experience of East Asia, but has also worked on Latin America, South Asia, Africa, Europe and Central Asia, and the OECD countries. He holds a DPhil in Economics from the University of York. Before joining the World Bank, he was Professor of Economics at the University of Sussex. He also held research positions at the universities of Aberdeen and York. Adam was Associate Editor of the *Journal of Health Economics* for 20 years, and is President Elect of the International Health Economics Association. Adam's research interests are in public economics and development economics, with health and distributional issues being his main preoccupation. Within health, he has published extensively on, inter alia, health financing and health insurance; provider payment reform; equality, equity and poverty; and the demand for and production of health. Outside the area of health, he has published on efficiency measurement, the redistributive effect and sources of progressivity of the personal income tax, and the redistributive effect of economic growth. He also has an interest and has published in the field of bibliometrics.

Homa Zarghamee is an Assistant Professor of Economics at Barnard College, the all-women's liberal arts college at Columbia University. She is a behavioral and experimental economist whose areas of research include emotions, subjective wellbeing, social preferences, and gender. She holds a PhD in Economics from Cornell University, with a focus on development economics.

Joseph Zeira is Professor of Economics at the Hebrew University of Jerusalem. He also has a part-time position in LUISS Guido Carli in Rome. He is a macroeconomist, who specializes in a number of areas: technology and economic growth, the role of income distribution in macroeconomics, money and liquidity, business cycles, and the economy of Israel. In addition to academic research and teaching he has been involved in a number of professional activities. He is an active member of AIX, a group of Israeli, Palestinian, and international scholars, who study the economic aspects of potential peace agreements between Israel and Palestine. He is also involved in many public debates on economic policy in Israel. In 2011 he headed a team of economists that supported the Israeli protest movement of 2011. In 2014 he was nominated to a public committee with the goal of reducing poverty in Israel.

Preface

Inequality and Growth: A Preamble[1]

Kaushik Basu, Joseph E. Stiglitz and Vivian Hon

It was a part of the common wisdom of mainstream economics that, for developing countries, in the early stages of development, inequality would rise but, as growth persisted, inequality would, eventually, decline. Evidence gathered over the first three decades after World War II, seemed to suggest that this pattern would be borne out. But, as more time passed and growth persisted, inequality, as measured on several dimensions, has continued to grow. What this illustrates is the folly of trying to determine long-run future trends by extrapolating from a couple of decades of data. Looking at the past is an uncertain guide for the future: the innovations created at each stage of history may or may not make the next fundamentally different from those of the past. To peer into the future, we will not only need data but also analysis and theory. Our aim in this volume is more modest than peering into the distant future; but rather, to analyze the current state of global and regional inequality, to dissect the phenomenal increase in inequality that we have seen occur in recent times, and to better understand the relationship between inequality and development. But taking a cue from what was argued above, we have been mindful to bring analysis and economic theory to bear on data and statistics. This was one of the driving forces behind the conceptualization of this monograph, which eventually grew to being a two-volume set. But there were other driving forces.

As the world has continued to grow, the persistence of extreme poverty and the growing gap in the incomes and wellbeing between the world's poorest and the richest people have become unconscionably high. As is argued in one of the chapters, there is growing evidence that there is not just a glass ceiling for the very poor but a glass floor for those who are born very rich. That is, it is difficult for them to become poor. We should care not only about the average growth rates within the economy, but also about how those numbers translate into opportunities for individuals – all individuals, both those born to the rich and to the poor. The long-run relationship between growth and inequality is important to study but we also have to try to understand the contemporary patterns and regularities, including those relating income inequalities and inequalities of opportunity. And as we acquire better knowledge of these, we have to ask ourselves what are the accompanying policy challenges for growth that is not just rapid but also inclusive and sustainable, which is nothing but

inclusive *over time*. Growth that benefits the current generation at the expense of future generations is not, in a fundamental sense, inclusive.

It is possible that the growing income and welfare gaps are contributing to the growing political turmoil we see around the world, from the Arab Spring and growing refugee crisis to the Occupy movement. Each one of these developments has distinct proximate causes, but it is arguable that underlying them there is a deeper cause rooted in the sense of despair and deep feelings of injustice and inequity and the inability to exercise voice through normal political processes for a mass of people who have felt their relative positions deteriorate. Extreme inequality not only deprives masses from sharing the benefits of economic development but by robbing the disadvantaged of voice, it also has a propensity to erode democracy.

Fortunately, as the challenge of inequality and marginalization has grown, several new books and writings have appeared, from Stiglitz (2012), Galbraith (2012), Piketty (2014), through Milanovic (2014) and Bourguignon (2015), to most recently, Atkinson (2015). Of late, this is a topic in which there has also been welcome engagement both from unlikely institutions such as the International Monetary Fund (Berg and Ostry, 2011) and the OECD (2014) and from less-surprising ones, such as Oxfam (2015).

In 2013, the World Bank Group officially declared two mission goals for itself – the end of chronic, extreme poverty by 2030 and the promotion of shared prosperity in every society, the latter being defined (for the purpose of statistical computation) as the promotion of growth of the per capita real income of the poorest 40 percent of each society.[2] Implicit in this was the mission to help the poorest segment of society to grow faster than the rest, thereby mitigating inequality. This was the first time that the Bank set the curbing of inequality as a mission goal.[3]

The importance of a focus on inequality stems from the fact that the bulk of inequality is not a matter of individual choice – some people preferring leisure to work and so choosing to be poor. Indeed, the bulk of human inequality is determined by birth – by what a child inherits and the kind of schooling, education, and health care the child receives. Since there are no hardworking babies, this inequality cannot be driven by individual preferences over leisure and work.

Once it is agreed that the right direction to go from where we stand today is to strive to mitigate poverty and inequality, attention must turn to the drivers of poverty and inequality. This is where the relation between growth and inequality becomes significant. Growth is of course very important, especially for low-income countries and emerging economies. It is not possible for such countries to have significant across-the-board reductions in poverty without growth. But growth does not necessarily lead to the reduction of poverty.

More broadly, what does growth mean for the reduction in inequality, and what does inequality do for growth? These questions have led to a contentious debate. There is an emerging consensus that inequality is bad for growth; and this is an idea that the reader will encounter in this two-volume book.

Some believers in "growth alone" (the view that the World Bank and other development agencies should focus just on growth) argue that growth will trickle down to the poor, unmindful of the fact that the word "trickle" itself gives away the hand of these commentators. This has been a contentious area of debate with a plethora of fallacies and misstatements. Consider, for instance, the finding, based on the study of large data sets, that over three quarters of poverty mitigation in recent decades can be attributed to plain and simple economic growth. From this, some people have deduced that growth therefore is the best cure for poverty and that we simply need to press on the growth accelerator and poverty will be taken care of.[4]

This, however, turns out to be a classic case of faulty deduction from arguably correct data. The fact that the bulk of the poverty eradicated in recent decades was because of growth does not mean growth is the best cure for poverty. All depends on what else was tried. If hardly any other relevant policies were tried, whatever poverty is eradicated would be because of growth, but that says little. It is like a Soviet economist studying job creation in the USSR in the 60s, 70s and 80s, and concluding that the government is the best creator of jobs, since almost all jobs were created by the state.

Indeed, even if growth by itself led to poverty reduction, given our poor record of eradicating poverty – in 2011, 14.5 percent of the world population lived below the poverty line of 1.25 dollars, Purchasing Power Parity (PPP)-adjusted),[5] per day – there is clearly a need to do much more and to look for appropriate policy interventions that go beyond "just" promoting growth.

This two-volume collection, *Inequality and Growth: Patterns and Policy,* is an effort to assemble the best of contemporary thinking on the subject. It is based on a roundtable convened by the International Economic Association (IEA) and the World Bank on "Shared Prosperity and Growth," and organized as part of the IEA 17th World Congress held at Dead Sea, Jordan, on June 10–11, 2014. In the roundtable that we organized, we tried to assemble an outstanding group of scholars, many of whom have grappled with the issues for many years, and from various perspectives.

The topics include conceptual issues and measurements, the state of global inequality, regional experiences, inequality of opportunity, consequences of inequality, and also some special areas that go beyond the traditional inequality discourse. The insights generated at the roundtable are critical in policy debates on economic development. The collection includes a total of sixteen full-length papers, and fifteen commentaries on those papers.[6]

The first volume, *Concepts and Analysis,* is a collection of papers on the conceptual and theoretical issues on inequality and its measurements. In chapter 1 of this volume entitled "New Theoretical Perspectives on the Distribution of Income and Wealth Among Individuals," Joseph Stiglitz lays out five new stylized facts. First, there is growing inequality in both wages and capital income (wealth) and growing inequality overall. Second, wealth is more unequally distributed than wages. Third, average wages have stagnated, even as productivity has increased, and so the share of capital has increased. Fourth, there have been significant increases in the wealth-income ratio, and last, the return to capital has not declined, even as the wealth-income ratio has increased.

Section 1 of this chapter provides an overview of the key anomalies presented by the new stylized facts. It explains why they are inconsistent with the standard neoclassical models long used by economists, and explains how a focus on rents (ignored in the standard neoclassical models) helps to resolve the inconsistencies. It explains the central confusion between "wealth" as a measure of control over resources and "capital" (or more broadly, an aggregate measure of productive capital), and shows that the former can be increasing while the latter is decreasing (at least relative to income or effective labor supply). Land values can increase, but the productive capacities of the economy decrease. An increase in wealth as a result of exploitation (monopoly) rents can even result in a decrease in productive capacities as measured wealth increases. For instance, savings data from the National Income Accounts for the United States account for only a fraction of the observed increase in wealth; the rest is associated with an increased in the capitalized value of rents. Section 2 of the first chapter re-examines the equilibrium wealth distribution within the context of a standard neoclassical model of growth without land, showing that (contrary to Piketty's assertions) there is not ever increasing wealth and income inequality. Equilibrium inequality is related to underlying behavioral and technology parameters. Section 3 introduces land and rents into the model, analyzes the long-run determinants of land rents and the price of land, and explains how the credit system plays an important role in both the increase in the wealth-income ratio and the increase in wealth inequality. The chapter explains that in designing policies to mitigate inequality, one has to be sensitive to the possibility of tax shifting, but shows how capital taxes with revenues devoted to investment as well as land taxes can lead to reduced inequality and higher steady-state income levels.

In the chapter "Reflections on the 'Equity and Development': World Development Report Ten Years Later," Francois Bourguignon reflects on the 2006 *World Development Report* (WDR). While the report represented a major step forward in the Bank's broadening its focus beyond poverty and growth as the main poverty reduction tool, it fell short in putting the issues of inequality front and center stage. Besides political considerations, this choice was dictated

by the recognition that the analytical link between inequality and development was more complex than a direct negative impact of income inequality on growth. Instead, the report focused on the inequality of opportunities as the major direct impediment to development and the ways of reducing it by "leveling the playing field", including among other measures of income redistribution. The increased attention to inequality of opportunity was a significant contribution to the development policy debate.

Ravi Kanbur and Adam Wagstaff's chapter, "How Useful is Inequality of Opportunity as a Policy Construct?" acknowledges that the empirical analysis of equality of opportunity has contributed significantly to our understanding of the determinants of inequality of outcomes and have become an important part of the analysis of policy. At the same time, as the authors point out, there are limitations, both conceptual and empirical, of the practical usefulness of the concept for policymaking. Drawing on applications in the education and health sector, they argue, as Francois Bourguignon does in the previous chapter, that the focus on inequality of opportunity is often used to delegitimize concerns over inequality of outcomes, including those arising from luck, risk and the distribution of talent (which is also luck, one step removed).

In "The Effects of Fiscal Redistribution," Michele Battisti and Joseph Zeira use cross-country, pooled, and panel regressions to examine the role of fiscal policy in reducing income inequality. They test for the possibility of any reverse causality, namely that public spending increase is driven by pressure to redistribute and which type of fiscal policy is most strongly related to the redistribution of income. They find that a one percent increase in public expenditure as a percent of GDP reduces income inequality in both cross-country and pooled regressions by 0.3–0.4 percent. Fiscal policy is also significant in reducing poverty, in particular labor market subsidies. Due to data constraints, their analysis on poverty is limited to OECD countries.

The papers by Hai-Anh Dang and Peter Lanjouw, Tony Castleman, James Foster and Stephen Smith, and by John Ifcher and Homa Zarghamee propose alternative ways of conceptualizing and measuring inequality, poverty and vulnerability. The simplicity of poverty headcount measures have made them the most widely used measure for monitoring poverty. However, in their chapter, "Person Equivalent Headcount Measures of Poverty," Tony Castleman, James Foster and Stephen Smith argue that the headcount measures ignore the intensity of poverty and this incentivizes policymakers to focus their efforts on the least deprived segments of the poor, since it costs the least to help this group to cross over the poverty line. They contend that other conceptually robust measures are often dismissed from the policy discourse as being too complex and not intuitive. The "person equivalent headcount measures" they propose uses a monetary, benchmark measure of the average depth of poverty to count up the number of "person equivalent" poor. They calculate the person equivalent

headcount for eighty countries using the World Bank's $1.25 per day poverty line, which shows a more rapid decline in global poverty and significant redistribution across regions and countries.[7]

In "Towards a Definition of Shared Prosperity: A Dynamic Perspective from Three Countries," Hai-Anh Dang and Peter Lanjouw present a complementary measure to the shared prosperity measure employed by the World Bank which considers not only the currently poor, but incorporates the vulnerable population, the segment of the population that is currently non-poor but face a heightened risk of falling back into poverty. Using illustrations from India, the United States and Vietnam for the mid- to late-2000s, they find that the two approaches are qualitatively consistent, with Vietnam enjoying the greatest boost in shared prosperity, followed by India and lastly, the United States.

John Ifcher and Homa Zarghamee's chapter on "Evidence of the Compression of the Subjective Wellbeing Distribution with Economic Growth" looks at inequality from the subjective wellbeing perspective. To date, the existing literature has focused mainly on the mean of subjective wellbeing. The chapter expands the body of work to looking at inequality of subjective wellbeing, its changes and the relationship to growth. Using data from the World Values Survey and World Development Indicators, Ifcher and Zarghamee find that per capita income is inversely related to subjective wellbeing inequality in cross-sectional and time series (excluding the two fastest growing economies). The latter is an interesting corollary to the "Easterlin Paradox" – that, though as income increases, happiness increases, in the long run, increased income is not correlated with increased happiness.

Karla Hoff broadens the inequality discourse by looking at social exclusion from the perspective of behavioral economics. In her paper, "Behavioral Economics and Social Exclusion: Can Interventions Overcome Prejudice?" she demonstrates that mental models – intuitive, socially learned sets of ideas about how things work – can bias an individual's perceptions of himself and the world. She argues that government programs should attempt to look at both structural and behavioral factors in addressing social exclusion. Group deliberation changed perceptions and overcame biases in ways that led to the abandonment of female genital mutilation in many villages in Senegal. In the West Bengal state in India, political affirmative action for women improved the way men perceived women, parents perceived their daughters, and women perceived themselves. However, political affirmative action for low castes (the Scheduled Castes) appear to have had no impact on broadly shared mental models. In the Indian state of Uttar Pradesh, affirmative action for Scheduled Caste did not appear to change the perceptions that the high-castes held of the Scheduled Castes and may have led to worse performance by high-caste teachers in public education, which Scheduled Castes unlike high castes disproportionately depend on. In India and China, experiments showed the

impact of activating existing mental models, rather than of trying to change them. Making salient (by publicly revelation) the social identities of students in dominant and stigmatized social groups created a gap between their average performance, with the stigmatized groups performing significantly worse.

The second volume, *Regions and Regularities,* examines the state of global inequality and inequality in different regions; and analyzes other kinds of inequality and discrimination.

In the first chapter, "The Inheritance of Employers and Nonlinearities in Intergenerational Earnings Mobility," Miles Corak and Patrizio Piraino examine intergenerational earnings mobility by looking at the role of parents on a child's interface with the labor market using a rich data set from Canada. They show that this explains nonlinearities in the intergenerational transmission of earnings. Getting a job in the father's firm plays a major role in preserving wealth and income advantages across generations.

In his chapter, "Do Nations Just Get the Inequality They Deserve?" José Gabriel Palma analyzes the contrasting centripetal and centrifugal forces at work within the distribution of income across countries. He argues that as a result of a process of convergence, the population located in the middle and upper-middle (i.e., within deciles 5 to 9) tends to appropriate a share of about 50 percent of the national income. As a result, he proposes an alternative inequality measure to the Gini which is often referred to as the "Palma ratio". The Palma ratio defined as the ratio of the income share of the top 10 percent over that of the bottom 40 percent, tries to capture inequality where it currently exists (the top and bottom of the income distribution). The chapter suggests that the huge diversity of distributional outcomes across the globe is not just the result of abstract economic forces, but rather the consequence of differences in economic structure and political settlements.

The chapter by Nora Lustig, Luis Felipe Lopez-Calva and Eduardo Ortiz-Juarez examines the state of inequality in the Latin America and the Caribbean region In "Deconstructing the Decline in Inequality in Latin America" they show that inequality, as measured by the Gini coefficient and other indicators (including all variations of the Kuznets ratio), declined in sixteen of the eighteen countries in Latin America and the Caribbean during the period 2000 to 2012. They attribute the decline to the decrease in hourly labor income inequality and progressive government transfers. This is a study that clearly has lessons for other parts of the world and other economies at similar stages of development.

The Arab countries are the focus of Radwan Shaban's chapter, "Inequality in Arab Countries." He observed a similar general declining trend in the Arab countries, in the period leading up to the Arab Spring. Furthermore, the median Gini coefficient for the Arab countries, at 36 percent, was lower compared to 38 percent for the world and 40 percent for all emerging market economies and

developing countries. He concludes by offering some plausible explanations of the difference between the measurement of inequality from household surveys and the perceived inequality as evidenced by the increased demand for fairness and social justice in the Arab countries.

In their chapter, James Galbraith, Beatrice Halbach, Aleksandra Malinowska, Amin Shams and Wenjie Zhang summarize a comprehensive revision and update of the University of Texas Inequality Project (UTIP) work on the inequality of pay and incomes around the world for the period 1963–2008. Their new data set on industrial pay inequality (UTIP-UNIDO) based on the Industrial Statistics of the United Nations Industrial Development Organization covers 4,054 observations for 167 countries, and the revised Estimated Household Income Inequality (EHII) database of gross household income inequality covers 3,871 observations for 149 countries. Their paper "The UTIP Global Inequality Data Sets 1963–2008: Updates, Revisions and Quality Checks" provides a fairly comprehensive quality check of the database against other available measures. They conclude that the EHII data set is reliable in estimating trends and reasonably reliable in estimating the level of gross income inequality observed in household surveys, but provides the advantage of dense and consistent coverage across the global economy, and is therefore useful for comparative and historical analyses.

In "Inequality and the Fragility of Growth," Jonathan Ostry attempts to address two questions. To what extent does inequality render growth more fragile? And, if inequality makes growth less stable, what are the possible implications as far as redistributive policies are concerned? Ostry examines the relationship between the duration of growth spells and a number of determinants, including inequality. He focuses on spells rather than standard panel growth regressions. He finds that more equal societies have more durable growth spells. He then asks the question whether redistributive policies should be used to bring about less fragile growth. Ostry finds that except in extreme cases, there is no trade-off between growth and redistribution – a marked departure from the perspective that was dominant until recently, but consistent with the analysis of Stiglitz (2012) and other more recent studies.

To what extent does inequality in the control over a society's resources facilitate or hinder growth? This is one of the key questions in the growth and inequality debate. However, empirical studies to date have tended to use the distribution of income as a proxy for distribution of wealth. In "Does Wealth Distribution and the Source of Wealth Matter for Economic Growth? Inherited v. Uninherited Billionaire Wealth and Billionaires' Political Connections," Sutirtha Bagchi and Jan Svejnar attempt to answer the question posed by the title of their paper, using a derived global measure of wealth

inequality from *Forbes* magazine's list of billionaire and decomposing wealth into three components: wealth obtained through political connections (cronyism), wealth generated from entrepreneurship, and inherited wealth. In their sample, inherited billionaires, the largest group account for about 54–72 percent of the total (depending on the year under consideration) and politically connected billionaires represent the smallest 4–13 percent of total billionaire wealth. They find that politically connected wealth and inherited wealth have a significant negative effect on growth, while the effects on growth of wealth generated from entrepreneurship are insignificant.

Ashwini Deshpande's chapter on "Caste Discrimination in Contemporary India" examines the state of caste disparities and discrimination in India, and notes how discrimination is very much a part of modern Indian society and also across the world with different kinds of systems, which seem to suggest that discrimination based on social identities is compatible with freely functioning markets. She finds that the persistence of caste inequalities results in both inequality of opportunities and inequality of outcomes. To rectify intergroup discrimination, such as those based on caste, will require purposive multi-pronged interventions.

The chapters are followed by interesting commentaries and in some cases, vigorous challenges by the discussants: Martin Ravallion, Bhaskar Dutta, Aristomene Varoudakis, Sudhir Anand, Paola Giuliano, James Foster and Murray Leibbrandt in volume 1 and Francisco Ferreira, Joseph Stiglitz, Edward Wolff, Francois Bourguignon, Kendra Bischoff, Shantayanan Devarajan, Celestin Monga, and William Darity, Jr. in volume 2. The work in these two volumes illustrates the complexity of the processes that determine the level of inequality in society and the importance of understanding them. The chapters have drawn attention to the many dimensions of inequality and the difficulties of measuring them, and the importance of both the inequality of outcomes and inequality of opportunity. The chapters have shown that inequalities, both in income and opportunity, can be affected by policy; and that policies that reduce these inequalities tend to promote development and enhance growth.

Some of the subjects addressed in this book have long been neglected. We see this two-volume monograph as opening up a debate, while being aware that there is much more to be said about each of the questions that have been addressed here. The aim of this volume is to put the discourse on a more sound scientific footing by marshalling some of the leading experts to contribute to our understanding of the main patterns of, and interconnections between, inequality and growth, and to nudge us towards the design of more effective policies for creating a better and a more inclusive global economy.

Notes

1. The authors would like to thank Karla Hoff for her extensive comments.
2. The twin goals were adopted by the World Bank's Governors at the Development Committee meeting on April 20, 2013.
3. It is worth clarifying that prior to the newly enunciated goals, many in the World Bank had written and spoken about the excesses of inequality; what is new is the formal adoption of the mitigation of inequality as a goal. Some commentators argued that the "bottom 40 percent" focus is not an inequality measure because it does not speak comparatively with the rest of the income distribution. While this is technically true, since the overall growth rate of each economy is so widely known, making the data available for the growth of the bottom 40 percent immediately shows whether the poorest 40 percent is catching up or falling behind. Making this data available is a prompt to policymakers to curb the growth of inequality, to make sure that the bottom 40 percent grows faster than the top 60 percent. To say that this is not an inequality measure because we do not say what the policymakers should do with the new data is like saying that providing the Gini coefficient is not an aid to curbing inequality unless we make it explicit each time that the aim is to minimize the Gini and not maximize it.
4. Here we overlook a number of complex statistical issues. If, as reflected in the emerging consensus, countries with lower inequality grow better and are more stable, and if (as is surely the case), lower inequality and greater stability are associated with lower poverty, there are difficult problems of identification that are not addressed by most of the statistical studies: policies that led to poverty and inequality reduction may have led to higher growth.
5. For those unfamiliar with purchasing power parity (PPP) numbers, we may point out that in most developing countries 1.25 dollars PPP-adjusted translates to 35 to 45 cents. The fact that nearly one-seventh of the world lives below this line is a serious indictment of our effort at poverty mitigation thus far.
6. Stiglitz's paper was planned as a general presentation for open discussion without a designated discussant.
7. Or $1.90 per day, which is the updated poverty line in 2011 based on new PPP data, but constant in real terms to $1.25 per day poverty line based on 2005 PPP data.

References

Atkinson, Anthony (2015) *Inequality: What Can be Done?* Boston: Harvard University Press.

Berg, Andrew and Jonathan D. Ostry (2011) "Inequality and Unsustainable Growth: Two Sides of the Same Coin?," IMF Staff Discussion Note SDN/11/08, April.

Bourguignon, Francois (2015) *The Globalization of Inequality*. Princeton, NJ: Princeton University Press.

Galbraith, James K. (2012) *Inequality and Instability: A Study of the World Economy Just Before the Great Crisis*. New York: Oxford University Press.

Milanovic, Branko (2013) "Global Inequality in Numbers: In History and Now," *Global Policy*, May.

OECD (2014) Focus on Inequality and Growth. www.oecd.org/social/inequality-and-poverty.htm.

Oxfam (2015) *Having it All and Wanting More*.

Piketty, Thomas (2014) *Capital in the Twenty-First Century*. Cambridge, MA: Harvard University Press. (First published as *Le Capital au XXI Siècle* in 2013).
Stiglitz, Joseph E. (2012) *The Price of Inequality: How Today's Divided Society Endangers Our Future*, New York: W.W. Norton.

1

New Theoretical Perspectives on the Distribution of Income and Wealth Among Individuals

Joseph E. Stiglitz[1]
Columbia University

1.1 New stylized facts of growth and distribution

A central question of economics has been: how do we explain the distribution of income among factors of production, and the distribution of income and wealth among individuals.

Some fifty years ago, theorists tried to develop explanations for what were then viewed to be the stylized facts of growth and distribution, articulated, for instance, by Nicholas Kaldor.[2] Among the central facts was the constancy of the capital–output ratio and the relative shares.

Today, there seems to be a new set of stylized facts that have to be explained, many of them markedly different from those that were the center of attention a half century ago.[3] Among the empirical observations are the following (some of these "facts" are truer for some countries than others; and there are a few country exceptions):[4]

(a) Growing inequality in both wages and capital income (wealth), and growing inequality overall.[5]
(b) Wealth is more unequally distributed than wages.
(c) Average wages have stagnated, even as productivity has increased, so the share of capital has increased.[6]
(d) Significant increases in the wealth–income ratio.[7]
(e) The return to capital has not declined, even as wealth–income ratio has increased.

The new stylized facts put a new light on Kuznets' hypothesis[8] that, while in earlier stages of development, inequality would grow, eventually inequality would fall. While that may have been true in the golden age of capitalism, between the end of World War II and around 1980, the period in which Kuznets was writing, such a conclusion no longer seems warranted.

In particular, Piketty (2014) has presented data showing that the decades following World War II were an historical anomaly, the one period in which capitalism was not characterized by a high level of inequality. He argues that not only has there been a large increase in inequality since 1980, but that the wealth of the economy, largely held by those at the top, will continue to grow faster than the overall economy.[9] If capitalists save all of their income, their wealth will grow at the rate of return, *r*, and if, as he hypothesizes, that is persistently above the rate of growth of the economy, *g*, their wealth relative to national income will grow at the rate of $r-g$.

Anecdotes aren't proofs, but they sometimes can alert us to factors that might have escaped attention in a simple model. John D. Rockefeller was America's first billionaire. At death, in 1937, his assets amounted to 1.5 percent of GDP. Had his assets grown at the rate "*g*" (the rate of growth of the economy) they would be worth today some $340 billion. If *r* (the relevant rate of return) were just 1 percent more than *g*, their family wealth should have grown to $680 billion. If, using numbers that Piketty might say are still conservative, but more realistic, the disparity between *g* and *r* is 2 percent, then their wealth would have been $1.3 trillion. Instead, the total value of the family assets is estimated to be $10 billion – less than 1 percent of the predicted amount – divided among almost 300 members.[10]

A critique

Three criticisms are raised against the Piketty analysis. First, once it is recognized that even capitalists consume, and that workers save out of wages (for life-cycle savings), then the neat relationship posited by Piketty for the ever-increasing capital–income ratio and inequality breaks down. For the wealth–income ratio of capitalists to be ever increasing would require $sr > g$, but in standard Solow model of growth, where workers save at the same rate that capitalists do, that inequality does not hold in the long run.

Secondly, the return to capital should be treated as endogenous. If the increase in wealth represented an increase in "capital," then the law of diminishing returns would imply that the return to capital should have decreased. Once account is taken of the endogeneity of r, a more subtle analysis of the determinants of wealth inequality is required. Indeed, even the central policy proposal, a (global) capital tax may not have the desired effect if there is tax shifting.

The disparity between *W* and *K* and the growth in land and other rents

Thirdly, and most importantly, while both wealth and capital are aggregates, they are distinctly different concepts. Once one recognizes this, it becomes easy to reconcile the stylized facts with conventional theory. The wealth–income ratio could be increasing even as the capital–income ratio (appropriately measured) is stagnating or decreasing. Much of wealth is not produced assets

("machines") but land[11] or other ownership claims giving rise to rents.[12] Some of the increase in wealth is the increase in the capitalized value of what might be called exploitation rents – associated with monopoly rents and rents arising from other deviations from the standard competitive paradigm. Some is an increase in the value of rents associated with intellectual property.

But that forces the analysis back one step: how do we explain the increase in the magnitudes of rents and the value of these assets? And what is the relationship between the increase in the value of these assets and the increase in inequality?

An analysis of the forces giving rise to the increase in land values and exploitation rents provides some insights into why there has been such a marked increase in wealth (and income) inequality, enables us to assess whether such increases are likely to continue, and to identify policies that might militate against these increases. If these assets are disproportionately owned by the rich, policies that lead to an increase in the value of these assets could have a first-order effect in increasing wealth inequality. We suggest that tax and financial market policies may have had these effects, and thus may have played an important role in the creation of today's high levels of inequality.[13]

Explaining the stylized facts

Solow, Kaldor, and a host of other economists produced a variety of models explaining the *old* stylized facts. But on the face of it, this would suggest that they cannot explain the markedly different *new* stylized facts. It would seem that a new set of theories is required.

This paper argues that only a slight (in the technical sense) modification of the old theories is required; but that while the modification may be technically small, this new theory has profound implications for how we view the economy, including for policy. Solow, and those working in the neoclassical tradition, assumed markets were competitive, and that output was produced with labor and capital, with a constant returns to scale production function. In that theory, rents played no role, because under those assumptions, there were no rents. We argue, however, that changes in rents, broadly defined – including land rents, exploitation rents, and rents on intellectual property – may be at the center of what has been happening; much of the increase in wealth is a result of the increase in (the capitalized value of) rents – and such increases do not increase, and may even decrease, economic output. Economic analysis should focus on how changes in technology (including innovations that may have enhanced the ability of those with market power to leverage that power), institutions, and policy may have increased these rents.

Equilibrium theories

This paper attempts to provide a set of coherent models that explain, or at least provide insights into, the new stylized facts. As in our earlier work,[14] a key part of our analysis is the insistence that there be consistency between the

micro-behavior of agents and the macro-behavior of the economy, and that crucial variables, like the rate of return on capital, be treated as *endogenous*.

The paper argues that we can best understand what has been happening as a shift from one equilibrium to another. Overall wealth inequality is related both to the transmission mechanisms for human and financial capital across generations and to life cycle savings.[15] In the models explored here, there is an equilibrium distribution between inherited and life-cycle savings; but changes in key parameters can change that equilibrium.

The models presented here differ, however, from earlier work in the analysis of income and wealth distribution in four ways: (a) We explicitly consider models in which there is a second, non-produced, asset, land; (b) We develop models in which while many individuals' saving is primarily for life-cycle purposes, there are a group of "capitalists" who pass significant amounts of wealth across generations; (c) We consider the possibility that the economy might not be fully competitive, and that there could be changes in the degree of market power; and (d) Land can be used as collateral, and the value of land (or other fixed assets) may be affected by financial and monetary policies.

The organization of the paper

The paper is distributed in four parts. Part I provides an overview of the key anomalies presented by the new stylized facts, and explains how a focus on rents helps to resolve them. Part II re-examines the equilibrium wealth distribution within the context of a standard model *without land*. Part III takes up the observation of Part I that a large proportion of the increase in wealth is related to the increase in the price of real estate. It was understandable why land was ignored in earlier neoclassical models (including Solow's, and those, like my own, trying to explain inequality): in a modern economy, land is not a central input into production. But this is not quite true. About a quarter of GDP represents housing services, of which land rents represent a significant proportion. (See the discussion below.)

It was the omission of land that represents the most important lacuna in my 1969 theory of the equilibrium distribution of wealth and income, which this paper attempts to rectify. We develop several models explaining the determination of the price of land, demonstrating why much of the increase in wealth would go into the value of land. It has long been recognized that there is a close link between financialization and inequality (Galbraith, 2012). We provide a set of models detailing that relationship, describing how when some assets are collateralizable and others are not, a change in financial/monetary policy can affect the value of collateralizable wealth. We explain why the composition of wealth between capitalists and life cycle savers are different; and financial and monetary policies that differentially affect different assets can have accordingly a marked effect on wealth distribution. More generally, we argue that the way

our credit system functions (or mal-functions) has played an important role both in the increase in the wealth–income ratio and in the increase in wealth inequality.

Part I: Key Anomalies and Their Resolution

The puzzles presented by the new stylized facts. As we noted in the introduction, economists had worked hard to explain the *old* stylized facts, and the theories they developed in response – and indeed theories developed over the past two hundred years – are challenged by the new stylized facts:

(i) The standard theories predict that the capital–labor ratio eventually is a constant. The new "theory" suggests that it is ever increasing (at a rate equal to $g-r$).
(ii) Standard growth theory begins with the observation that r, the rate of return on capital, is an *endogenous* variable. Among the most basic laws of economics is the law of diminishing returns. If capitalists continue to invest at a rate faster than the growth of the labor force,[16] then the rate of return to capital should diminish.[17]
(iii) Standard theories suggest that if the capital–output ratio increases, it is because there has been an increase in the capital–labor ratio.[18] An increase in the capital–labor ratio should be associated not only with a decrease in the return to capital r but as with an increase in wages; but as we have noted, wages have stagnated.
(iv) And while most (but not all) studies of the elasticity of substitution suggest that it is less than unity, capital deepening would imply an increasing share of labor – contrary to the new stylized facts.[19,20]
(v) It is hard to reconcile the increase in the wealth–income ratio with national income account data on savings. There is a large unexplained component, which we call the *wealth* (or wealth-income) *residual*.

It is thus hard to reconcile several of the new stylized facts with standard neoclassical theory, if we interpret wealth, W, in the usual way as capital, K. In the first two subsections, we elaborate on these puzzles, providing the resolution in section 2.3.

1.2 Key anomalies and their resolution

1.2.1 The wealth–accumulation residual

Here, we focus on the last of the puzzles: how we can reconcile the magnitude of the increase in wealth (capital) with national accounting data on savings.

Just as a matter of national accounting, if s is the fraction of national income saved (net),

$$\frac{d}{dt}(\log K) \equiv \frac{sY}{K} \tag{2.1}$$

And

$$\frac{d}{dt}\left(\log\left(\frac{K}{Y}\right)\right) \equiv \frac{sY}{K} - g. \tag{2.2}$$

Piketty and Zucman present data showing that the average net national savings rate of the US over the period 1970–2010 is 5.2 percent,[21] and that the average growth rate of the economy was 2.8 percent. The wealth–income ratio varied, beginning the period at just under 4 and ending at about 4.6. Thus, treating for the moment "*K*" and "*W*" as identical (2.2) would have predicted a decline in the wealth–income ratio, at an average annual rate of somewhat more than 1.5 percent, in contrast to the observed increase. If these numbers were accurate, the observed increase in wealth income ratios must come from somewhere else than the steady accumulation of capital goods.[22]

This can be thought of as the "wealth-accumulation residual" (analogous to the Solow residual – Solow had shown that capital accumulation could account for only a small fraction of the increase in productivity; the rest had to be explained *somehow.*) We will argue below that there is a simple explanation of the residual – the increase in the capitalized value of rents, including land rents.

We can reframe (2.2) to ask, what is the critical net savings rate such that there is an increase in the "real" capital–output ratio? Let k be the effective capital–labor ratio, g^* be the "natural" rate of growth of the economy, the sum of the rate of growth of population (work force) and the rate of labor augmenting technological progress, $\beta = W/Y$, and $\xi = K/W$, the ratio of the value of produced capital to wealth (which includes land); then

$$\frac{d}{dt}\left(\log\left(\frac{K}{Y}\right)\right) = \frac{sY}{K} - g^* = \frac{s}{\beta\xi} - g^*, \tag{2.3}$$

so that capital deepening (defined as an increase in the capital output ratio) occurs if and only if

$$s > g^*\beta\xi. \tag{2.4}$$

If it were assumed that the US growth over the last forty years was close to its natural rate, 2.8 percent, $\beta = 4$, and $\xi = 1$ (land is an unimportant), then s would have to be greater than 11.2 percent, more than twice the net savings

rate for the US. More realistic, even if $\xi=.8$, s would have to be greater than 8.9 percent. Given the US savings rate of 5.2 percent, only if $\xi<.46$ will there be capital deepening.

The US is an open economy, and there have been considerable capital inflows. These have varied considerably at a percentage of GDP. Assume capital inflows equal iY. Then

$$d \log (K/Y)/dt=(s+i)Y/K-g^*=(s+i)/\beta\xi-g^*. \quad (2.3')$$

Thus, adding to the earlier parameters ($\beta=4$; $\xi=1$; $s=5.2$ percent) a reasonable value of $i\approx.02$, d log $(K/Y)/dt\approx-1$ percent. Even taking account of capital inflows, the capital–output ratio falls at the rate of about 1 percent per year.[23]

Even if the savings rates were slightly higher, or the return to capital slightly higher, it is hard to generate plausible increases in the *real* capital stock that could account for the observed increases in the *wealth–income* ratios in recent decades.[24]

There is still a different way of looking at the puzzle of the increase in wealth–output ratios. Over the past sixty years, a wide variety of models describing the growth of the economy have been formulated. In each, in the long run (steady state) there is a particular capital–output ratio. In each, changes in the underlying parameters (the rate of growth of the labor force, the rate of growth of labor-augmenting technological progress, and savings behavior) can explain a change in the long run capital–output ratio. The question is, have there been any changes in these parameters sufficient to explain/account for changes in the capital–output ratio and the factor distribution of income of the magnitude observed?

For instance, in the Solow growth model, the long-run capital–output ratio is given by s/g^*, where again g^* is the long-run growth rate, equal to the rate of growth of labor supply plus labor- augmenting technological change, and s is the savings rate.[25] g^* has varied, for instance increasing in the 1990s and the first part of this century, while the savings rate (in the US) has decreased, which would suggest a decrease in the long-un capital–output ratio, not an increase – let alone an increase of the magnitude asserted.[26,27]

1.2.2 Can wages fall, the capital–output ratio increase, and the return to capital not fall as k increases?

The previous section argued that in none of the standard models of economic growth can one plausibly obtain an increase in the equilibrium value of the capital–output ratio of the magnitude observed *if we interpret wealth as capital*. If one interprets "*W*" as capital, then there has been not only an increase in the capital–output ratio, but also in the capital–labor ratio. Our ultimate objective is to understand the distribution of income, both among individuals and

among factor shares. We now ask, can wages fall (as they have been) as k (the capital–labor ratio) increases, *within the standard neoclassical model*.

Movements in average wages. Some have suggested that some forms of capital are like robots, and compete directly with workers, lowering their wages. But highly skilled workers still need to manage the robots, and even if the increased capital lowers the return to unskilled workers, it increases the return to the skilled workers. In Appendix A we show that under standard assumptions, an appropriately weighted average wage *must* increase.

Data for the United States, for instance, shows otherwise: a stagnating or declining average wage rate during the past four decades, during which the capital–output ratio has increased – if we interpret "wealth" as capital.[28]

Movements in average productivity. Unfortunately, we typically cannot observe marginal productivities directly; but we do have data on average productivities, $\bar{P}=F(K)/L$. It should be obvious that (if nothing else changes) $\frac{d\bar{P}}{dK} = \frac{F_K}{L} > 0$, i.e. average productivity should increase with capital deepening.

Direct data on average productivity is consistent with this hypothesis. Thus, if we are to believe in the competitive determination of wages, given the large disparity in the movement of, say, the average productivity of the bottom 99 percent and their average wage, then somehow a huge gap between movements in marginal and average productivities must have opened up – a gap that has yet to be explained.[29]

Technological change. There is a related hypothesis: that technological change has diminished the returns to unskilled labor. It is skill biased.[30] While the timing of the changes in the share of labor and the decrease even in wages of relatively skilled labor in more recent years argues against skill biased technological change as the major or at least sole explanation of changes in distribution,[31] here we focus on the analytics.

If there were a single type of labor, then labor–augmenting technological change increases the effective labor supply, and, everything else being the same, would reduce the effective capital–labor ratio, and hence the wage *per effective labor unit*. But each worker would represent a larger number of effective labor units, so whether the wage per worker increases or decreases would depend on the elasticity of substitution.[32] Only if the elasticity of substitution is substantially below unity would wages fall. (As we noted earlier, interpreting wealth as "K" implies an elasticity of substitution *greater* than unity, which would imply an increase in wages. Similar results hold in the longer run, when there is an adjustment in the capital stock.[33])

Assume now there are two types of labor, skilled and unskilled, and technology is skilled biased, say increasing the productivity of the skilled workers, while leaving that of unskilled workers unchanged. Whatever the factor bias of technological change, it must move the factor price frontier outwards,

which means that if the return to capital doesn't change, then the return to at least one of the two types of labor must increase. It is possible to show that if the return to capital remains unchanged, the average wage would have to increase.[34] Again, it is not easy to reconcile observed patterns of changes in factor prices with the theory.[35]

1.2.3 The resolution of the seeming paradox: There is more to wealth than capital

The previous two sections argued that it is hard to reconcile the new stylized facts with virtually any form of the standard growth model *under the assumption that the increase in wealth corresponds to an increase in productive capital*. What then is going on?

The most plausible hypothesis is that wealth (W) and capital (K) are markedly different objects (as Piketty himself recognizes, but the full implications of which he does not take on board), and that wealth can be going up even as capital (as conventionally understood) is going down. If capital is not going up much (or even going down) in tandem with the increase in the effective labor supply, it would explain why the interest rate has not gone down. (As we note below, we need to go further to explain the failure of the average wage to rise.)

There are many forms of wealth that are not produced assets. Much of the increase in wealth in recent years is associated with an increase in the value of land. The increase in the value of land does not, however, mean that there is *more* land, and that therefore the productivity of labor should go up. And an increase in the value of land does not mean that the marginal productivity of capital should decrease. Once we sever the relationship between K and W, all the paradoxes described in the previous section disappear.

Wealth as a measure of control over resources. The standard wealth income measure, constructed by adding up the money value of wealth and dividing it by the money value of income. Tracing how that ratio, and ownership of that wealth, evolves over time captures something that is important in our economy and how it is changing: control over resources. But changes in the wealth distribution, so measured, do not even necessarily reflect well the distribution of "wellbeing." For the bundles of goods bought by those at different income/wealth levels may differ – indeed, in some of the models below, the increase in wealth is closely linked to the increase in the price of a good which is consumed only by the rich, so that the increase in inequality in wellbeing is markedly lower than the increase in money-wealth.[36]

But what is clear is that the measure of wealth so constructed is *not* a good measure of the relevant inputs into the production process – wealth could be going up, and yet any reasonable measure of inputs could be moving in the opposite direction.

Index number problems and wealth as a measure of productive inputs.[37] Not only are the concepts different, but there are difficult measurement problems involved in each. Both are aggregates, and an aggregate constructed for one purpose may not be appropriate for another. The "volume" of capital goods resulting from saving out of national income (letting consumption goods be the numeraire) will be affected by changes in the price of capital goods relative to consumption goods. And the *effective* increase in "*K*" will also be affected by capital augmenting technological change. (Indeed, the two issues are closely related; because there are constant changes in the design of capital goods, one has to establish a "hedonic" index of equivalency.) If the only capital good were computers, the increase in the "volume" of *K* from a given amount of savings would have increased enormously over time. In calculating aggregate "*K*," we have to add up capital of different types, whose relative prices and productivities are changing over time.

But even abstracting from these subtleties, and assuming that there were a single capital good, *K*, and a single fixed factor, land, *T*, we can easily see that movements in K do not adequately summarize what is happening to aggregate input (relative to labor). If land is a factor of production,[38] then wages will be related to inputs of both *K* and *T*. If *T* is fixed, then the increase in *K* has to be proportionally greater–possibly much greater – than the increase in labor supply to ensure that wages increase, to offset the failure of *T* to rise.

In short, we need to add up *K* and *T* somehow to ascertain what is happening to the aggregate input, which we will refer to as ***C***. How we add the two together matters a great deal. And what makes sense for one purpose or in the context of one model or an economy with one technology may not in another.

If *T* and *K* were additive in the production function i.e. $Y=F(K+T, L)$, then to assess what is happening to the aggregate input, which we call ***C***, we simply add *K* and *T* up linearly.[39] In the case of France, this aggregate "***C***" has been going up more slowly than GDP, even though *K* has been going up *slightly* faster than GDP (see Figure 1.1).[40]

On the other hand, we could have a production function of the form

$$Y=F(\boldsymbol{C}, L) \tag{2.5}$$

where now

$$\boldsymbol{C}=K^{\zeta}T^{1-\zeta}. \tag{2.6}$$

Then, since *T* is fixed,

$$\frac{d}{dt}(\log(\boldsymbol{C})) = \zeta \frac{d}{dt}(\log(K)). \tag{2.7}$$

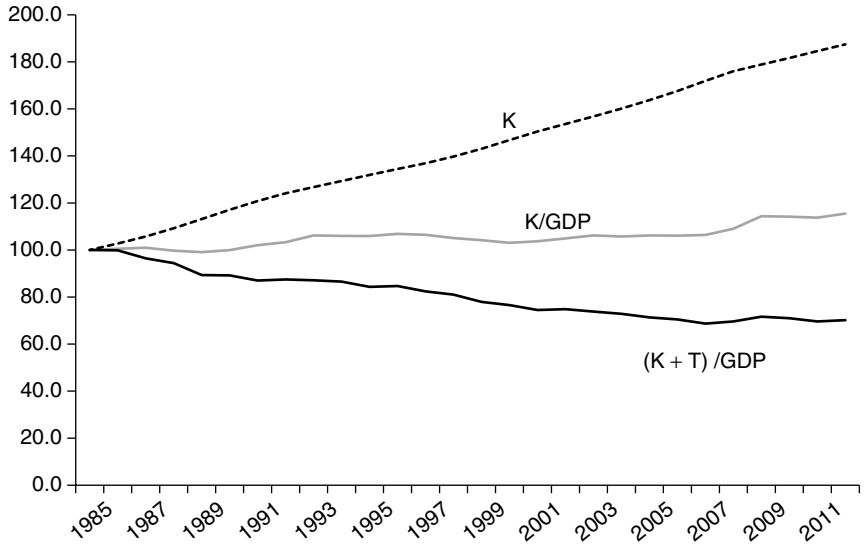


Figure 1.1 GDP and inputs in France
Source: Chart provided by Paul Schreyer, based on OECD national accounts data and INSEE Comptes du patrimoine.

Now, $\boldsymbol{C}$ is increasing if K is increasing, but whether it is increasing faster or slower than GDP depends on the relative weights assigned to the two inputs, ζ. With even a relatively high value of ζ, $\boldsymbol{C}/Y$ appears to be declining for France.

Notice that for the United States, d log "$\boldsymbol{C}$"/dln t$\approx$.01 ζ<.028, so that even if the wealth–income ratio is increasing, $\frac{\boldsymbol{C}}{Y}$ is declining at a rapid rate, in excess of 1 percent per year.

The production function defined by (2.5) and (2.6) has the interesting property that W increases in proportion to K, but it would be totally wrong to confuse W with K. More generally, depending on the parameter ζ, the rate of increase in W can be much larger or smaller than that in K. $\frac{d(W/Y)}{dt} > 0$ while $\frac{d}{dt}\left(\frac{\boldsymbol{C}}{Y}\right) < 0$ if (refer to Appendix B for a formal treatment)

$$g\beta\xi < s < g\beta. \tag{2.8}$$

As we noted, for the United States, the latter inequality is clearly satisfied, while for plausibly small values of ξ, so is the former.

This analysis makes clear that different indices, different measures of $\boldsymbol{C}$, can differ not just in the magnitude by which they change over time, but even in the direction of change; and an appropriate measure of aggregate input could have gone down even though the standard measure of wealth increased.

Other data problems. This section has explained why data on wealth do not reflect "capital." Several of the stylized facts involved inequality metrics. There are serious problems associated with measuring the factor distribution. Because our tax system taxes capital gains at a lower rate than ordinary wage income, there are incentives to try to recategorize labor income as capital income (for example, private equity and carried interest). Going the other way, large fractions of the income of banks is paid out in bonuses to their managers, and thus treated as wage income in the national accounts. Likewise for the managers in other corporations. But there is a fundamental difference between these payments and ordinary wages. To a large extent, the managers determine their own pay. Though often referred to as incentive pay, the link between pay and performance is weak, evidenced so clearly in the 2008 recession;[41] the money can better be thought of as a return on the control rights of the firm. While such property rights normally are not sold or bought in open markets (though occasionally they are, often with much contestation), they are transferred from one group of managers to their successors, and in the process there can be a significant gift exchange (that is, a provision of even a more generous retirement benefit than was contracted for) in the expectation of a similar transfer upon their retirement. If we appropriately relabel such income as non-wage income, then the share of wages would have declined even more than shown by the standard data series.[42,43]

1.2.4 Parsing out the wealth residual

We argued in section 1.2.1 that it is hard to reconcile national savings data with the observed increase in wealth. There was what we referred to as the "wealth residual." There are, in fact, three reasons that W can increase without a concomitant increase in *K*, besides an increase in the value of land. There could be an increase in the value of other inelastically supplied factors.[44] There can be an increase in the value of intellectual property. Or there can be an increase in what might be called "exploitation" rents. In the discussion below, we will use the term "market power" and "exploitation" interchangeably. The deviations from the competitive benchmark that we are interested in here take on many forms besides that classically associated with imperfect competition in product or labor markets. There can also be exploitation by corporate or other special interests of the public: indeed, it was in this context that the term rent-seeking first got coined.

Some of the increase in wealth, as we shall see, has as much to do with our accounting frameworks as with anything else. Some of these instances of an increase in measured wealth are actually associated with decreases in the effective productivity of the economy.

Changes in rents on land and other non-produced assets. In later sections of this paper we model the determination of land rents and the value of fixed

assets. A decrease in the interest rate (normally associated with capital deepening) should lead to an increase in the value of such assets. As population increases, the scarcity value of particularly attractive sites (like land in the Riviera) becomes greater. Much of the value of land today is in urban areas; as the population in key urban centers increases,[45] the value of land in these cities increases.

There is considerable evidence that recent decades have shown "a historically unprecedented boom in global house prices... Rising land prices explain about 80 percent of the global house price boom that has taken place since World War II."[46] The increase in land prices thus accounts for much of the increase in wealth and wealth–income ratios.

There can be an increase in the value of *any* asset fixed in supply: The wealthy strive not just to own homes in the Riviera but also Renaissance paintings. Thus, the discussion of positional goods in Part IV of this paper applies to these other assets as well as to land. In a world with increasing population, and fixed supplies of depletable natural resources, the value of these resources too can be expected to increase.[47]

Changes in market power and exploitation. There is an increasing consensus that much of observed inequality – especially at the top – is associated with rent seeking, including the exercise of monopoly power.[48] If monopoly power of firms increases, it will show up as an increase in the income of capital, and the present discounted value of that will show up as an increase in wealth (since claims on the rents associated with that market power can be bought and sold.)[49]

The magnitude of the associated increases in the capital–wealth ratio from even a small increase in exploitation can be significant. A permanent increase in the share of capital by just 1 percent would, when capitalized at a real discount rate of 1.5 percent, imply an increase of the wealth–income ratio of .67; an increase of market exploitation leading to an increase in the share of capital by 5 percent would lead to an increase in the wealth–income ratio by more than 3.[50]

There is an extensive literature discussing why we might expect an increase in monopoly power in a modern economy, for example, as a result of network externalities (Katz and Shapiro, 1994) and the fixed costs associated with research (Dasgupta and Stiglitz, 1980). (Many of these arguments, however, are inconsistent with the assumption of a constant returns to scale production function.) So too, the transformation of the economy towards the service sectors may have increased the importance of local monopolies (see Greenwald and Kahn, 2009). Note that such increases in wealth are associated with a decrease in the economy's effective productivity, because they are associated with an increase in market distortions. Moreover, it is an implication of such exploitation that even though W is increasing, wages are decreasing.

While increases in monopoly rents are the most obvious example of an increase in wealth not associated with an increase in the productive capacity of the economy, there are many other forms of exploitation which may have increased in recent decades; the capitalized value of any such change would show up as a change in wealth.

Elsewhere, we and others (Galbraith, 2012) have focused on the role of the financial sector in increasing inequality. The financial sector grew before the 2008 crisis from 2 percent to 8 percent of GDP. Profits grew to absorbing 40 percent of all corporate profits. There are reasons to believe that much of this might be associated with exploitation rents (including those associated with market manipulation, insider trading, predatory lending,[51] and anti-competitive practices arising from their control of the payments mechanisms, giving rise as well to abusive practices in credit and debit cards, and so on) and capitalized in the value of wealth. Though there was some increase in the amount of wealth to be managed, the increase in the wealth–income ratio was not so substantial to account for the increase in the share of the financial sector; nor can that sector's remuneration be accounted for by the improvements in their management of the funds, and even less so, by any improvement in overall economic performance.[52]

If the financial sector improved its ability to exploit the poor through predatory and discriminatory lending practices and abusive credit card practices (and the resulting profits were not bid away because of imperfections of competition) then there would be an increase in standard metrics of wealth.[53]

Other forms of exploitation of consumers. The financial sector has perhaps deservedly earned a reputation for its ability to exploit – to take advantage of imperfections of information and limitations of individuals' ability to process information. But other sectors have also increased their capacity to create and exploit such imperfections. Behavioral economics has exposed a large number of "irrationalities" in individuals' behavior, instances for example in which individuals systematically overestimate some risk and underestimate others. Corporations have now begun systematically to exploit such irrationalities to increase their profits.

Successful corporate rent-seeking: transfers from the public sector to the private. There are more subtle forms of "exploitation." Government allows "too-big-to-fail" banks. The value of those banks is higher than they otherwise would be, *because of government risk-absorption*. But the contingent liability of the government is not capitalized, and because this liability doesn't show up in the national balance sheet, it appears as if the wealth of the economy has increased. But with appropriate metrics (where the decreased wealth of wage-earning citizens, as a result of the increase in the expected present discounted value of the higher taxes that they will have to pay to bail out the banks), just the opposite would have happened: we would have recognized that because of

the distortions associated with too-big-to-fail banks, the productive capacity of the economy has been diminished; that the bail-outs are Pareto-inefficient, and that the wealth of the economy has been diminished.[54]

In each of these situations, a change in the flow of resources that accrues to "capital" gets capitalized in wealth, and the present discounted value of the decreased flow to the rest of the economy is not reflected in our wealth metrics. *We don't, for instance, value the change in the stream of tax revenues to the government or the expenditures by the government or the reduced wages accruing to workers as a result of increased market exploitation.*

Knowledge and information rents. Earlier, we explained how firms can generate rents by creating and exploiting information asymmetries. In a modern economy, there are many other ways by which knowledge and information differentials can give rise to rents. Insider trading and market manipulation (e.g. in the Libor and Foreign Exchange markets) are the most obvious examples. There are reasons to believe that much of the profits generated by high frequency trading is a sophisticated form of front-running, taking advantage of differential access to information (Stiglitz, 2014c). These information rents are often primarily distributive, increasing incomes of some individuals at the expense of others. In some cases, they even lead to Pareto inefficiency.[55] When capitalized, however, they lead to an increase in wealth, even if net income is decreased.

Intellectual property. There is another, closely related and increasingly important category of assets, intellectual property. Here, there have been three factors contributing to the increased market value of intellectual property: there may be more knowledge; the value of any "piece" of knowledge increases as the size of the economy (other inputs) increase – knowledge and these other inputs are complementary; and more of knowledge has been privately appropriated, and hence shows up in wealth data.[56] Knowledge that is freely available increases output, but doesn't show up in anybody's balance sheet and therefore would not normally be reflected in the national accounts as wealth. But changes in the intellectual property regime (what Boyle (2003) refers to as the enclosure of the knowledge commons) has resulted in an increase in the wealth of those who are given these property rights.[57]

Changes in discount rates and risk management. There is a further reason for an increase in the value of wealth without a concomitant increase in the *physical* productive capital stock: the rate of discount may fall – for example, because of a decrease in the interest rate – and this may induce large changes in the relative price of different goods (and in the price of capital goods relative to consumption). This was the essential issue in the Cambridge–Cambridge controversy some half a century ago, where it was observed that the value of capital and the choice of technique may be non-monotonic in the interest rate.[58]

In the private sector, the relevant discount rate is the after tax return, so that there are two offsetting effects on the value of wealth of an increase in

the tax on capital. In the limiting case where before tax returns are unaffected, the value of an asset yielding a before-tax return of R every year would be unchanged i.e. $V = \frac{(1-t)R}{(1-t)r} = R/r.$ The value of assets facing an average tax rate greater than that relevant for the discount rate will go down; and conversely if the average tax rate is smaller.

Changes in risk management and the ability to absorb risk can also have effects on the wealth–income ratio.[59] At the same mean and variance of the return to an asset, such changes lead to an increase in the certainty equivalent return, and therefore of the market value. If the improved risk management/ability to absorb risk leads to a lower discount rate, the increase in market value can be even larger.

There can also be countervailing general equilibrium effects. Individuals may reallocate more of their wealth to assets with a higher risk and higher mean return, i.e. assets which (on average) have a *lower* capital–income ratio.

Part II: Equilibrium Wealth Distributions in Neoclassical Models

A key concern in the growing inequality in the United States and other advanced countries is the worry that we are giving rise to an inherited plutocracy. Piketty (2014) emphasized that if $s_p=1$ and the rate of interest were greater than the rate of growth, inherited wealth would increase faster than the growth in income. On the other hand, the fact that individuals are living longer and must save for their retirement means that life cycle savings is increasing, reflected in part in the huge increase in pension funds.[60] In this section, we construct a simple model incorporating both inherited and life cycle savings.

We are able to obtain simple formulae describing the equilibrium share of wealth held by life cycle savers. Using these formulae, we can easily ascertain the effects of, say, tax policy or changes in the parameters of the economy. We show that an increase in the savings rate of workers (as a result, for instance of encouraging them to save more) has no effect on output per capita, but does increase the share of wealth of life cycle savers. Life cycle savings crowds out inherited savings. On the other hand, a tax on capital (even if it is paid disproportionately by the rich capitalists, with proceeds paid out to workers, and so is therefore viewed as progressive) will be so shifted that capitalists are unaffected and workers' income, including transfers, actually goes down, as does their share in national wealth. This bears out a general theme of this paper: tax policies have to be constructed to take into account general equilibrium incidence effects.

1.3 Savings models

This section is divided into two parts. The first presents the basic model, while in the second, we assume all individuals have identical savings functions. The only difference is that when wealth is low enough, bequests drop to zero.

1.3.1 Basic Model

We assume two groups: There are workers who live two periods, and save for their retirement.[61] Their savings is referred to as "life cycle savings." Then there are the capitalists, who save a fixed percentage of their income, s_p.[62] For simplicity, we use a discrete time model.

In this section, output is produced by means of a neoclassical constant returns to scale production function Q=F(K,L), where K is the capital stock and L the labor supply (there is full employment). k=K/L is the capital–labor ratio. Q/L=F/L=f(k) gives output per worker as a function of the capital–labor ratio. The return to capital is f′, and the wage rate is f−kf′. We assume that the number of capitalists and workers increase at the same rate, n (assumed here to be exogenous.) (In this simple version, we ignore labor augmenting technological progress. It is straightforward to bring it into the analysis.)

The difference equations describing the evolution of the system are given by[63]

$$(1+n)k^c_{t+1}=(1+s_pf'(k_t))k^c_t \tag{3.1}$$

and

$$k^w_{t+1}=(1+r)\,s(k_{t+1})w(k_t)/(1+\mathrm{n}) \tag{3.2}$$

where k^w and k^c are workers' and capitalists' capital (per capita), respectively, where we have allowed the savings rate of workers to depend on the (rationally expected) interest rate,[64] and where

$$k_t=k^w_t+\beta k^c_t, \tag{3.3}$$

where β is the ratio of workers to capitalists. (By assumption capitalists supply no labor. Recall that k, which enters the production function, is the ratio of the capital stock to *workers*, not the per capita capital stock.) β is assumed to be fixed.

These equations fully describe the dynamics, given an initial value of workers' and capitalists' capital.[65] In the steady state, $k^{c*}=k^c_t=k^c_{t+1}$ and similarly for k^w_t. Hence, from (3.1)

$$n=s_pf'(k^*), \tag{3.4}$$

where k^* is the steady-state value of k and $f'(k^*)$ is the steady-state return on capital, equal to r. Note that r here is the return over a generation, i.e. if a generation is 30 years, and the annual interest rate is 2 percent, $r \approx 1$. The steady-state level of capital (and the equilibrium interest rate) is determined simply by capitalists' saving propensity.

If workers save more, the economy does not become richer; income does not go up; wages do not increase. All that happens is that they increase their share of total capital.

The steady-state capital of workers (life cycle capital) given by (where we look at workers' wealth as of the beginning of their period of retirement, after earning interest, rather than at the end of their working life)

$$k^{w*} = (1 + r^*)\, s(k^*) w(k^*)/1 + \mathrm{n} \tag{3.5}$$

Hence

$$\frac{k^{w*}}{k^*} = \frac{s(k^*)w(k^*)}{(1+n)k^*}(1+r^*) \tag{3.6}$$

Using (3.4) this can be rewritten

$$\frac{k^{w*}}{k^*} = \frac{s(k^*)w(k^*)\left(1+\frac{n}{s_p}\right)}{(1+n)f'^{-1}\left(\frac{n}{s_p}\right)} = n\frac{s(k^*)\left(1+\frac{n}{s_p}\right)}{(1+n)s_p}\frac{1-S_k}{S_k} \tag{3.7}$$

The ratio of wealth of life-cycle savers to that of capitalists (or to total wealth) depends on the relative savings rates, the relative shares (recall that S_k is the share of capital), and the growth rate. A decrease in the growth rate would (if the elasticity of substitution is less than one and if the savings rate did not change) lead to an increase in the capital–labor ratio and a decrease in the share of capital. There is a critical value of the elasticity of substitution, such that below that threshold, a decrease in the growth rate leads to an increased share of life-cycle savings, and above that threshold, it leads to a decreased share. (The rate of return to capital does not enter into this formula, because it is an endogenous variable. But this analysis has ignored the effects on workers' savings rate. A decrease in the growth rate leads to a lower interest rate, and this can lead to either a higher or lower value of s depending on the sign of s'.)[66]

If the savings rate of workers increases, for instance because of increased expected retirement longevity,[67] workers' wealth increases proportionately, while aggregate wealth remains unchanged. By the same token, in this model, if the

generosity of social security increases, so the savings rate of workers decreases, workers' wealth (excluding their claims on social security) decreases proportionately, while aggregate wealth remains unchanged (in a pay-as-you-go system).

There is an important qualification to this analysis: workers' savings has to be low enough so that, on their own, they do not drive the rate of return below n/s_p. For if they do, then the life cycle savers eventually drive out the capitalists.[68] It would appear that this condition is normally satisfied.

Market distortions The analysis so far has assumed that workers are able to get the same return on their investments as capitalists. We then obtain

$$\frac{k^{\wedge w*}}{k^{\wedge *}} = n\frac{s^*(1+\left(\frac{n}{s_p}\right)\varsigma(1-\tau^{cw})/(1-\tau^{cc}))}{s_p(1+n)}\frac{1-S_k}{S_k}$$

where r^w, the return workers receive on their investments, is ς times that of capitalists, and τ^{cw} is the effective tax rate on the return to capital for life cycle savings , τ^{cc} that on the return to capital for capitalists. Thus $\frac{k^{\wedge w*}}{k^{\wedge *}}$ will be lower than suggested by the basic model if (a) a distorted financial market delivers to life cycle savers lower returns than those received by capitalists; and (b) regressive taxation leads to life cycle savers facing higher tax rates (than those confronting capitalists). An example of the former that has recently been exposed is how conflicts of interest among those managing large fractions of IRA accounts lead to substantially lower returns on those accounts. Part II provided several other reasons for why life cycle savers might receive lower returns on their investments than do capitalists. The share of life cycle savings will be further lowered if, as we suggested in section 2, because of monopolies and other distortions the share of capital is larger than it would have been in a competitive equilibrium.

1.3.1.1 The effect of taxation

If we impose a tax on capital at the rate τ^c, we obtain instead of (3.4)

$$n=(1-\tau^c)s_p f'(k^*), \tag{3.4a}$$

implying that *the after-tax return to capital is not affected by the tax* (just as was the case in the Kaldor model). There is, in effect, full "shifting." As the tax rate increases, the equilibrium capital stock diminishes.[69]

Capital taxation with proceeds distributed to workers. To ascertain the effect on the relative importance of lifecycle savings, we have to specify what happens

to the tax revenue. Assume it is redistributed to workers. Then the transfer Ţ (per capita) is given by

$$Ţ = \tau^c r(k^*)k^*. \tag{3.8}$$

Noting that in our simplified model, the saving rate depends only on the after tax rate of return, and from (3.4a) that is unchanged, and letting s* denoted that value of s, (3.6) can be rewritten as

$$\frac{k^{w*}}{k^*} = \frac{\left(1+\frac{n}{s_p}\right)s^*(w(k^*)+)}{(1+n)k^*} \tag{3.9}$$

Then, to ascertain the effect of an increase in the tax rate on the share of inherited wealth, we simply have to ascertain the sign of

$$\frac{d\left(\frac{s^*(w(k^*)+Ţ)}{k^*}\right)}{d\tau^c}. \tag{3.10}$$

Normally, an increase in the tax rate lowers the wage, but at least for low τ^c increases the transfer.

Workers' lifetime income $Y^W = w(k^*)+Ţ$, so that[70]

$$\frac{dY^W}{d\tau^c} = (-k^*f''(k^*) + \tau^c(k^*f''(k^*) + f'(k^*)))\frac{dk^*}{d\tau^c} + r(k^*)k^* \tag{3.11}$$

Where

$$\frac{dk^*}{d\tau^c} = \frac{f'(k^*)}{(1-\tau^c)f''(k^*)}. \tag{3.12}$$

The sign of (3.11) is thus that of $\frac{\tau^c(f'(k^*))^2}{(1-\tau^c)f''(k^*)} < 0$ for $0<\tau^c<1$. ($\frac{dY^W}{d\tau^c} = 0$ at $\tau^c=0$.)

Hence, *the loss in wages is always greater than the benefit from the transfer.*

It follows that an increase in the interest income tax always increases the relative importance of inherited wealth.[71]

The tax also has an adverse effect on the distribution of consumption (wellbeing). Since the after-tax interest rate facing capitalists is the same, their flow of consumption (in steady state) is unaffected. Workers' lifetime utility is a function of their income, Y^W, and the interest they receive on their savings (after tax). We have already shown the derivative of Y^W with respect to τ^c is negative (except at $\tau^c=0$, where it is zero). But because the after-tax return the

worker receives from his investment is unaffected, workers are unambiguously worse off.

Thus, in the case that would *seem* to be the most favorable to workers – where all the proceeds are redistributed to them – their income is reduced, their welfare is reduced, and inequality is increased.

Inheritance tax with proceeds distributed to workers. With an inheritance tax, there is still tax shifting: wages fall and the before-tax return on capital for capitalists increases. Appendix C shows that the relative share of life cycle savings may increase, so long as the elasticity of substitution is not too small, and that there is an optimal tax rate, maximizing workers' wellbeing.

Public investment. So far, the results of this section on the ability of the government to improve the wealth distribution through capital taxation are somewhat disheartening. If, instead, government invests the tax proceeds as well as the proceeds it gets from its investments, then an increasing fraction of the capital stock will be owned by the government. The government investment drives down the return to capital, so that the wealth of the capitalists can't keep up with the increase in population. Their wealth diminishes (per capita), and we get a new equilibrium which is similar to the original equilibrium except that now the government owns all the capital and, in effect, its saving rate is unity. Then wages are higher, and workers are unambiguously better off. Note that this would be true even if the government were slightly less efficient than the private sector.[72]

If we expand the model to a three-factor production function, $Y=F(K_p, K_g, L)$, with private and public capital goods, and (some of) the proceeds from the tax are invested into the public capital good, then it is easy to show that there can be a new equilibrium in which a (somewhat poorer) capitalist class survives but the tax *may* still have a positive effect on workers: In a three-factor production function, K_p and L can be substitutes, and K_g and L can be complements, so that on both accounts, wages are increased as a result of the tax; but the increase in K_g is consistent with the after tax return to capital returning to its previous level.[73]

Progressive capital taxation[74] A progressive capital income tax can affect the degree of inequality among the rich.[75] The argument for a progressive capital tax is strengthened if we look more carefully at the nature of the *measured* returns to capital. In economists' simplest models, all capital receives the same returns. If returns are stochastic, then it is simply luck that determines who gets high returns. If that were all that there were to the matter, a progressive tax on the rate of return to capital in excess of the average return (with offsets for returns below that level) would be welfare increasing, if capitalists were risk averse. If savings were elastic in the certainty equivalent return, then savings would increase, and workers would be better off.

There may, however, be other possible explanations for above average returns. The returns could represent greater skill at investing, in which the returns

ought to be viewed as a return to labor, not as a return to capital.[76] The returns could represent a return to risk taking. If capital markets are imperfect (so risk is not fully diversified) and individuals are risk averse, riskier investments will yield higher returns than safe. A proportional capital tax on *excess returns* (over the safe rate of interest) would, under these circumstances, increase risk taking, and thereby average incomes. Finally, the returns could in part be a return to exploitation. To the extent that that is the case,[77] then a progressive tax would discourage such rent seeking behavior, increase economic efficiency, improve the wellbeing of those who are being exploited, and reduce overall inequality.

1.3.2 Toward a more general model

The previous subsection assumed that society is composed of two groups of individuals – workers who engage in life cycle savings, and capitalists who pass on wealth from one generation to the other. In fact, however, all individuals could have the same savings function; it is simply past circumstances that determine the observed savings rate. Assume, for instance, that providing bequests is a "luxury," and that when individuals' wealth exceeds a certain level, they begin to act like capitalists, passing on money to their heirs.

We assume gross savings of any individual are a function of his end of period wealth, which is just his wage and the return on the capital from the previous period: $s(W_t)W_t$, where

$$W_t = w_t + (1 + r_t)k_t \tag{3.13}$$

But assume $s(W_t)$ is S-shaped, the extreme version of which would be $s = s_0$ for $W \leq W^*$ and $s = s_1 \gg s_0$ for $W > W^*$.[78] Then there exists a two-class equilibrium. To see the nature of the equilibrium, assume a fixed fraction of the population ϑ are in the upper income group. Then

$$s_i(w(k)) + (1 + r(k)k_i) = (1 + n)\ (k_i)\ i = 0,1 \tag{3.14}$$

$$k = \vartheta k_1 + (1 - \vartheta)k_0 \tag{3.15}$$

For each value of ϑ, there is a different equilibrium, that is, $k_i = k_i(\vartheta)$.

Special cases of this model yield the standard Solow and Kaldor/Pasinetti/life cycle model. If $\vartheta = 0$, we obtain the discrete variant of the Solow model. On the other hand, if $s_0 \approx 0$, (3.14) can be approximated by

$$s_1\left(1 + r(k) + \frac{\vartheta}{k} w(k)\right) = 1 + n, \tag{3.16}$$

Here, it is not that the workers have a different savings *function* from that of the capitalists; it is only that their income is low so they save little. Most

importantly, we have endogenously *derived* a two-class model out of a S-shaped savings function.

In this model ϑ is determined just by history. For each, there is a steady state (k_1, k_2). Individuals never leave the "class" into which they are born. But it is easy to construct a stochastic model in which some in the upper class have bad luck and move down, and some in the lower have good luck and move up. ϑ is then solved for endogenously, related to the transition probabilities (see Stiglitz, 2015b). Changes in policy, behavior and technology (the savings functions, the stochastic processes) can move the economy from one in which most individuals are in the "upper group" (the middle class society of the past) to one in which most are in the lower group (the "99 percent/1 percent society of the present.) Financial sector "innovations" that encouraged those at lower wealth not to save and regressive capital taxation might, for instance, accomplish this.

Part III: Land Rents

In section I of this paper, we noted that standard neoclassical models focusing on capital and labor in competitive markets could not explain the increase in the wealth–output ratio observed in the US and many other advanced countries and other stylized facts of modern economies.[79] Central to our resolution of these puzzles, we suggested, was the understanding that wealth and capital were different concepts. The most important source of the disparity between the growth of wealth and the growth of productive capital is the growth of the value of land – not associated with any increase in the *amount* of land and therefore of the productivity of the economy.[80]

In this part, we present a series of models that might account for much of the increase in the value of wealth taking the form of an increase in the price of land. These models not only help us understand the increase in the wealth–income ratio, but also the increase in wealth inequality. This part is divided into five sections. In section 1.4, we extend the life cycle/inheritance model of section 1.3 to land. Section 1.5 presents the simplest model with land rents, showing that even in this very simple model, the increase in wealth may be markedly greater than the increase in capital. Section 5 examines land as a positional good, deriving a similar result that increases in wealth are greater than increases in capital. Section 1.6 investigates the dynamics of land prices, showing that in a natural formulation, bubbles can easily arise, and along such "bubble paths," wealth may increase, even though capital (per capita) is decreasing. In effect, wealth accumulation in the form of land may crowd out real capital accumulation.[81] The final section explores how financial and monetary policies can give rise to an increase in land prices and thus "wealth," but such increases in wealth may have little to do with what is happening to

the *real* wealth of the economy – which in this simple model is reflected in the value of the capital stock (per capita.)

There is one further (important) explanation of an increase in land values: the increase in urbanization leads to an increase in urban land values, the value of being in proximity to urban centers.[82]

1.4 Land in a life cycle model

In section 1.3, we formulated a life cycle model, and used it to explain the division of wealth between capitalists and workers (life time savers). It is easy to incorporate land into this framework. Now, however, because land is a store of value that is alternative to capital, there is an important question: could savings that otherwise be used for capital accumulation be deflected into land, thereby *harming* workers?

1.4.1 Pure life cycle model

We begin our analysis with the case where there are *only* life cycle savers, but there is a fixed asset, which we will call land.

It is useful to rewrite (3.2) to focus on "savings in capital":

$$s(w(k), r(k))w(k) - \frac{f_{T^\wedge}}{f_k} = k. \tag{4.1a}$$

Any value of k solving (4.1a) is a steady-state equilibrium.

There can be multiple equilibria, as illustrated in Figure 1.2. As k increases, wages increase. The slope of the LHS can be greater or less than unity, and can vary with k, so that the LHS can cross the 45 degree line more than once. There is a natural sense in which stability requires that the savings curve cut the 45 degree locus from above, i.e. the increase in savings *into capital* from an increase in the capital stock is less than the increase in the capital stock itself.

Looking across (steady-state) equilibria, it is clear that, letting W denote wealth per capita.

$$\frac{dW}{dk} = \frac{d}{dk}\left(k + \frac{f_{T^\wedge}}{f_k}\right) = 1 + f_{T^\wedge k}/f_k - f_{T^\wedge} f_{kk}/f_k^2. \tag{4.2}$$

If

$$f_{T^\wedge k}/f_k - f_{T^\wedge} f_{kk}/f_k^2 > 0, \tag{4.3}$$

then W increases more than k. That will always be the case if $T^\wedge$ and k are complements.

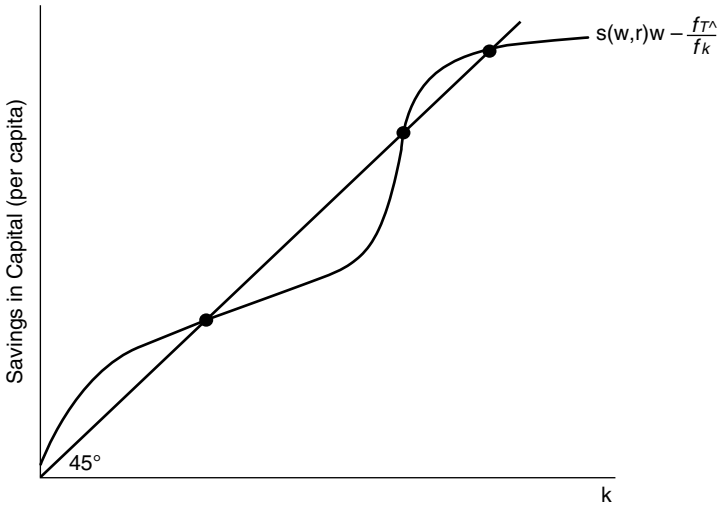


Figure 1.2 Multiple equilibria in pure life cycle mode

By the same token, we can ask what happens if there is an upward shift in the savings function, i.e. the savings function is given by $\gamma s(w(k), r(k))$. Then

$$\frac{dk}{d\gamma} = \frac{sw}{(1+f_{T^\wedge k}/f_k - f_{T^\wedge}f_{kk}/f_k^2)(1+n) - \gamma sw' - \gamma w\dfrac{ds}{dk}} \tag{4.4}$$

while, from (4.2),

$$\frac{dW}{d\gamma} = \frac{dW}{dk}\frac{dk}{d\gamma} = (1+f_{T^\wedge k}/f_k - f_T\, f_{kk}/f_k^2)\frac{dk}{d\gamma}. \tag{4.5}$$

Again, we get the result that W can increase more than k. Some of the increased savings goes into an increased value of land, reducing the benefits that otherwise would have accrued to a higher savings rate.

Taxing capital. A tax on the return to wealth (both land and capital) will shift the function $sw - f_{T^\wedge}/f_k$ up or down depending on whether s is decreasing or increasing in r (increasing or decreasing in k), which implies that in a stable equilibrium, it will lead to an increased or decreased value of k depending on whether s' is greater or less than zero. The change in wealth will typically be larger than the change in k (so long as inequality (4.3) is satisfied). But while in a two-factor production function, a decrease in k necessarily leads to a lower wage, now it may not. Capital and labor may be substitutes rather than complements. (Robots may be a substitute for unskilled labor.)

Taxing land. It is easy to see that in this model, a tax on the value of land the proceeds of which are distributed to workers results in an increase in investment and a reduction in the return to capital (in a stable equilibrium).[83] If $F_{KL} > 0$ (labor and capital are complements) wages will rise. *A fortiori,* if the revenues are fully invested, wages go up even more.

1.4.2 A two-class model

In this section, we return to our two-class model of section 1.3, but introduce land. For simplicity, we focus only on the steady state.[84] But this poses a problem in the absence of land-augmenting technological change and population growth: if the equilibrium interest rate would go to zero (as it would if n were equal to zero), the value of land would go to infinity. There are at least two ways out of this puzzle: (a) assume land does not yield any return or (b) assume land augmenting technological progress at the rate n. Here, we take the latter tack, and express all units in per capita terms (per unit of effective land).

The variables of interest can all be expressed as functions of k. The returns to land must equal the returns to holding capital. In steady state, the price of a unit of effective land, denoted by q, will be constant. Letting $f_{T^\wedge}$ denote the marginal return of a unit of effective land, which in steady state is constant, $q = \frac{f_{T^\wedge}}{f_k}$, in the obvious notation, where wages and returns to capital are functions of the capital stock per capita. Savings are put either into capital goods or into land holdings.

Instead of (3.1) the capitalists' wealth accumulation equation is described by

$$k^c_{t+1} + q_{t+1}T^c_{t+1} = W^c_{t+1} = \frac{k^c_t + q_t T^c_t + s_p f_k(k_t)(k^c_t + q_t T^c_t)}{1+n} = \frac{(1 + s_p f_k(k_t))W^c_t}{1+n} \qquad (4.6)$$

where T^c_t is the effective landholdings of the capitalists at time t (here, per capita) and q is the price of an effective unit of land. In steady state, the return to capital and the return to land (the return to each of the assets) is the same. The rate of interest must be equal to the rate of growth divided by the savings propensity of capitalists, as before, and that implies a particular value of $k=k^*$. We similarly rewrite (3.2) as (continuing with the obvious notation)

$$W^w_{(t+1)} = s(k_{(t+1)})w^w_t/1+n. \qquad (4.7)$$

Hence, the steady-state equation for life cycle wealth relative to total wealth is now just

$$\frac{W^{w*}}{W^*} = \frac{s(k^*)w(k^*)}{(1+n)W^*} = \frac{s(k^*)w(k^*)}{k^*\ (1+n)\ W^*}\ k^* = n\frac{s(k^*)}{(1+n)s_p}\frac{1-S_k}{S_k(1+x)}. \qquad (4.8)$$

where $\varkappa \equiv$ the ratio of the value of land to capital. In this case, $q^* = \frac{f_{T^\wedge}}{f_k - n}$. Changes in worker savings have no effect on wealth; an increase in capitalists' savings rate leads to an increase in *k*, with an effect on wealth that is normally greater than the increase in k because of the increased value of land, as in the earlier model.

We can easily study the effect of various forms of taxation on the distribution of income and wealth (between capitalists and life-cycle savers); these effects are markedly different than in the pure life cycle model of the previous subsection because of tax shifting. Land taxation has no effect on k^*, hence no effect on wages; it leads to a diminution of the value of wealth. If the proceeds of the tax are distributed to workers, life cycle wealth is increased, and therefore on both accounts, wealth inequality is reduced. (Similar results hold for land capital gains taxes.) Inheritance taxation, as in section 3, leads to an increase in the before tax return on capital, lowering k. If capital and labor are substitutes, then capital and land have to be complements, and the tax on inherited capital unambiguously reduces wealth inequality. Wages go up and the return to land goes down, so the share of wealth held in life cycle savings unambiguously goes up. But if capital and labor are complements, the opposite may happen.[85]

1.5 A simple model with land rents

To see more clearly the relationship between wealth and capital, we can formulate an even simpler model than the life cycle model of the previous section. Assume the rents associated with land are fixed and last in perpetuity, while the production of industrial goods requires no land. Then a slight decrease in the (long-term real) interest rate can lead to a large increase in the value of land.[86] Thus, national output is given by

$$Q=F(K,L)+R \tag{5.1}$$

where Q is total output, *K* is productive capital and *L* is labor, for the moment assumed fixed, F is constant returns to scale, and R is the fixed return to land. Then the value of wealth, *W*, is given by[87]

$$W = K + \frac{R}{r} = K + R/F_K, \tag{5.2}$$

where r is the rate of interest (return on capital, equal to F_K) so that

$$\frac{dW}{dK} = 1 - \frac{RF_{KK}}{F_K^2} > 1 \tag{5.3}$$

If F is, for instance, a unitary elasticity of substitution production function, with coefficient on capital of α, then

$$\frac{dW}{dK} = 1 + \frac{R}{Q-R}\frac{1-\alpha}{\alpha} \tag{5.4}$$

If, for instance, $R/Q = .3$ and $\alpha = .2$, then $dW/dK \approx 1 + 1.7 = 2.7$ the increase in wealth is more than twice the increase in the productive capital.

The effect of taxation. If the return to land is taxed, then W and K are more closely aligned. If the returns to land are fully taxed (as they would be with the Henry George tax), W and K would be fully aligned. This follows directly from rewriting (5.2) as

$$W = K + \frac{(1-\tau^L)R}{r} = K + (1-\tau^L)R/F_K, \tag{5.2'}$$

where τ^L is the tax rate on the returns to land.

1.6 Positional goods

Similarly, if land serves as a positional good, there can be an increase in the value of land, without any increase in the productive potential of the economy. Rich individuals compete for houses in the Riviera. As the rich get richer, they compete more vigorously for this real estate, and the price of this fixed asset increases, without any increase in "real" output.

Assume there are some assets in fixed supply (positional goods) that do not affect production of conventional goods. Assume all the wealth of the economy is held by the rich (an assumption which does not depart too far from reality) and that the demand by rich for these goods is given by $M(W, p)$ with the equilibrium given by

$$M(W, p) = pT \tag{6.1}$$

where p is price of land, T, which is fixed supply, and $W = K + pT$. For simplicity, we choose units so $T = 1$. (6.1) can be solved for p as a function of W, and K can then be solved for

$$K = W - p(W) \tag{6.2}$$

Then

$$\frac{dK}{dW} = 1 - p' = 1 - \frac{M_w}{1-M_p} < 1 \tag{6.3}$$

If the wealth elasticity of the demand for positional goods is large enough and the price elasticity is small enough, then an increase in W may even be associated with a decrease in K.

The effect of land taxation. As in the previous section, land taxation (and in more dynamic models, the taxation of capital gains on land) can help align K and W. The demand for positional goods depends not just (or even so much) on the price as on the "user cost" or opportunity cost: **M**(W, p, u). The opportunity cost is r, the return on capital. If there is a land tax, the cost of owning the positional good becomes $r+\tau^L p$. (In more general dynamic models, where the value of land is increasing, the user cost is $u = r + p\,[\tau^L - (1-\tau^{cg})\dfrac{dlog\ p}{dt}$, where τ^{cg} is the tax rate on capital gains.) Instead of (6.1) we have (6.1′) $\mathbf{M}\left(\mathrm{W},\ \mathrm{p},\ r + p\left[\tau^L - (1-\tau^{cg})\dfrac{dlog\ p}{dt}\right]\right) = \mathrm{p}$ which for fixed expectations about capital gains and zero capital gains tax rates can be solved for p as a function of K, W, and τ^L We can then rewrite (6.2) as

$$K = W - P(W, K, \tau^L). \tag{6.2′}$$

At any given K, the higher τ^L, the lower wealth: the tax reduces the gap between wealth and capital.[88]

Inequality in wellbeing. While in this and other models in this section, the increase in wealth may be largely (or entirely) due to an increase in land values, one might ask: does this lead to *real* inequality. After all, the rich consume the positional goods. The increase in land values affects them, and them only. Workers are only affected to the extent that the increase in land values crowds out capital accumulation, so K decreases (or does not increase as much as it otherwise would.)

While this conclusion is true in the simplified model we have constructed here, it is natural that there be a spill over to workers (and in practice, such spillovers typically occur.) Assume, for instance, landlords/capitalists rent out some of their land to workers, at a rental price of pF_K. Then, policies and behavior which lead to an increase in pF_K disadvantage workers.

Still, the observation that the increase in land prices (or of other positional goods) disproportionately affects the wealthy has several important implications. First, it reminds that in making comparisons across different income groups, we have to take into account the different market baskets of goods that they consume. The increase in the relative prices of positional goods means that there may not have been as large an increase in inequality as would appear to be the case.[89]

Secondly, it helps explain differences in savings behavior both over time and across income levels. To achieve "success" as demonstrated by acquiring expensive positional goods may require more savings (more wealth) today than

when the price of such goods were lower. It may be that there is a difference between savings out of capital gains, especially those arising from the increase in the value of real estate, and other returns to capital, precisely because of the consequences of those price changes for acquiring the goods in the future that the rich seek to purchase.

Thirdly, by the same token, patterns of inheritances and life-time giving across generations too may be endogenous, affected in particular by such changes. If increases in real estate prices make it difficult for even reasonably successful workers to purchase a home that they and their parents believe is appropriate to their station in life, wealthy parents will provide larger intra vivo transfers. Note that, in some sense, the direction of causality has changed: greater wealth and wealth inequality arising from an increase in real estate prices has led to greater inheritances and intra vivo transfers across generations among the top.[90]

Foreign ownership. The demand by foreigners for positional goods may lead to an increase in the wealth of the citizens of a country as well as to an increase in wealth inequality. Assume, as above, rentiers own all the positional goods (land in the Riviera). A sudden and unanticipated increase in the desire for these pieces of land by foreigners increases their value, and the wealth of those who happened to own this land; and if those within the country are the wealthy, it will contribute to the increase in inequality within the country. (This seems to have been a factor increasing inequality within several countries.)

1.7 Bubbles: the dynamic instability of the market economy

Bubbles are a pervasive and recurrent aspect of market economies. While the recession may have represented a "correction," the economy may not have fully corrected the price of real estate.[91]

Hahn and Shell-Stiglitz[92] showed the dynamic instability of the economy with heterogeneous capital goods in the absence of a full set of futures markets extending *infinitely* far into the future (or without perfect foresight extending infinitely far into the future). The steady state was a saddle point.

The same result also holds for a model with capital and land (with two state variables, K, the stock of capital, and p, the price of land). We extend the production function in the straightforward way so that $Y=F(K,L,T)$, where, as before, T is the supply of land and L is the supply of labor, and F is constant returns to scale.[93]

There is a delicate problem: without growth of the labor force, the equilibrium interest rate will be zero in the long run in the Kaldor model.[94] But at a zero interest rate, if there are positive returns to land, the value of land becomes infinite – in effect, the model breaks down. Assuming labor growth (or labor augmenting technological progress) poses its own problems: the land–labor ratio goes to zero, and under normal assumptions about the production

function, the return to land itself would go off to infinity. This problem can in turn be "solved" by assuming just the right amount of land-augmenting technological progress. At first blush, this seems unpersuasive: why should nature produce land-augmenting technological progress in just the right amount to sustain a steady state. But upon reflection, it may not be so coincidental, once we introduce a theory of endogenous factor bias. We know that the bias is determined by relative shares, and if the elasticity of substitution is less than one, as land becomes more scarce, there are greater incentives for land-augmenting technological progress.[95]

We investigate two alternative approaches. The first entails assuming a conventional production function (without land), but the existence of land as a store of value. The second assumes a fixed rate of land augmenting technological change, equal to n.

1.7.1 Non-productive land[96]

The key equilibrium condition is that the return to holding land and capital must be the same, i.e. since land is non-productive, its entire return is its capital gain, $\frac{d}{dt}(\log p)$, and the equilibrium condition is

$$\frac{d}{dt}(\log p) = (F_K - \mu) \tag{7.1}$$

where μ is the depreciation rate and F_K is the gross return to capital.

The short-run dynamics are described by (7.1) and

$$\frac{dK}{dt} + T\frac{dp}{dt} = s\left(F_K K - \mu K + T\frac{dp}{dt}\right) \tag{7.2}$$

where we have assumed that only capitalists save and they save a fixed fraction, s, of "full net income" including capital gains (Shell, Sidrauski, and Stiglitz, 1969).[97] The RHS of (7.2) is net savings (as seen by the individual, not according to the national income accounts). This goes into an increase in the value of land ("land savings") or capital accumulation.

Substituting (7.1) into (7.2), we obtain (again using the normalization that $T=1$):

$$\frac{dK}{dt} = s(F_K K - \mu K) - (1-s)\frac{dp}{dt} = (sK - (1-s)p)(F_K - \mu) \tag{7.3}$$

(7.3) and (7.1) provide a pair of differential equations fully describing the dynamics of the economy.

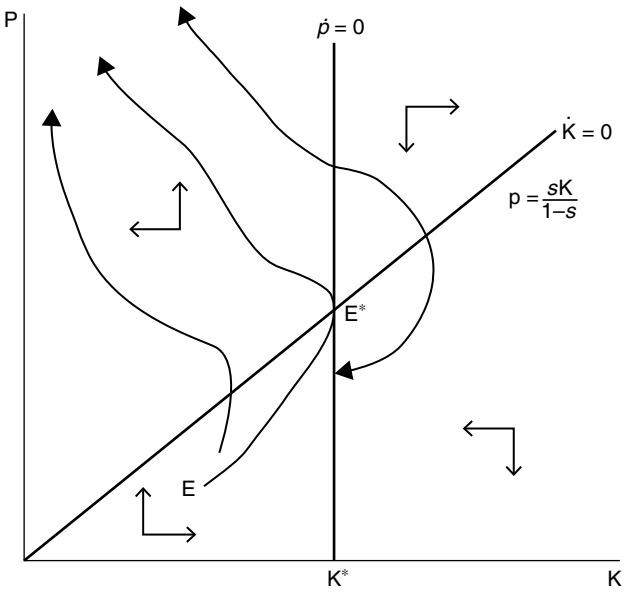


Figure 1.3 Steady states and dynamics with non-productive land: linear savings

Figure 1.3 shows the steady states, given by the solution to the loci

$$F_K = \mu \tag{7.4a}$$

And

$$p = sK/(1-s). \tag{7.4b}$$

We define K^* as the value of K solving (7.4a). Note that any value of p along $K=K^*$ is an equilibrium, since $dK/dt=0$ when $K=K^*$ (net income of capitalists is zero).

The dynamics are easy to describe and are also depicted in Figure 1.3: To the right of $K=K^*$, p is decreasing (the net return to capital is negative) and to the left it is increasing. Above the $dK/dt=0$ locus, but to the left of K^*, K is decreasing, while above the $dK/dt=0$ locus, to the right of K^*, K is increasing. Conversely, below the $dK/dt=0$ locus, but to the left of K^*, K is increasing, while below the $dK/dt=0$ locus, to the right of K^*, K is decreasing.

Let $p^* \equiv sK^*/(1-s)$. K^* in combination of any value of $p<p^*$ is a stable equilibrium; K^* and any value of $p \geq p^*$ is an unstable equilibrium. The saddle point trajectory EE^* divides the bottom quadrant (below $dK/dt=0$ and to the left

of K^*) into a convergent and non-convergent region. Below EE^*, paths converge to $K=K^*$. Above EE^*, they diverge. As a trajectory below the dK/dt locus and to the left of K^* approaches K^*, the slope is

$$\frac{dp}{dK} \approx \frac{p}{sK-(1-s)p} \tag{7.5}$$

which is finite below the locus $p=sK/(1-s)$. Hence, trajectories hit the vertical axis, at which point they remain in the steady state. We can similarly show that if $K_0>K^*$, K will also hit K; but if the initial value of $p>sK/(1-s)$, K will initially increase, before decreasing to K^*.

Thus, there are an infinity of stable equilibria, in all of which the level of income is the same, but in which there can be markedly different values of wealth ($K+pT$). pT is in this sense fully indeterminate. But if $K<K^*$ and the initial price is too high, the economy experiences a bubble.

A generalized savings function. These results are partly a consequence of the special savings function employed. More generally, we assume

$$\frac{dK}{dt}+\frac{dp}{dt}=s\left(K,p,\frac{dp}{dt}\right), \tag{7.6}$$

Net savings are a function of capital, the value of land, and capital gains. K and p affect savings both because they increase the income and wealth of the individual. This formulation recognizes, however, that aggregate savings may differ depending on the composition of wealth (i.e. it is not necessarily just a function of $K+pT$, aggregate wealth). This may be because the risk properties of these assets differ or the individuals who own these assets differ.

With this formulation, the dynamics are described by (7.1) and

$$\frac{dK}{dt}=s\left(K,p,\frac{dp}{dt}\right)-\frac{dp}{dt}=s(K,p,p(F_K-\mu))-p(F_K-\mu). \tag{7.7}$$

There are two possible (sets of) steady states. One is given by the solution to (7.4a) and[98]

$$s(K^*,p^*,0)=0. \tag{7.8}$$

If we assume (at $dp/dt=0$), $s_K>0$ and $s_p>0$ (in the absence of capital gains, an increase in wealth of any form leads to increased savings), then (at least near $K=K^*$) the $dK/dt=0$ is downward sloping. The dynamics are unstable (Figure 1.4a), and may be oscillatory, as illustrated in Figure 1.4b.[99] Even though the local dynamics are unstable, there may be a limit cycle. In particular,

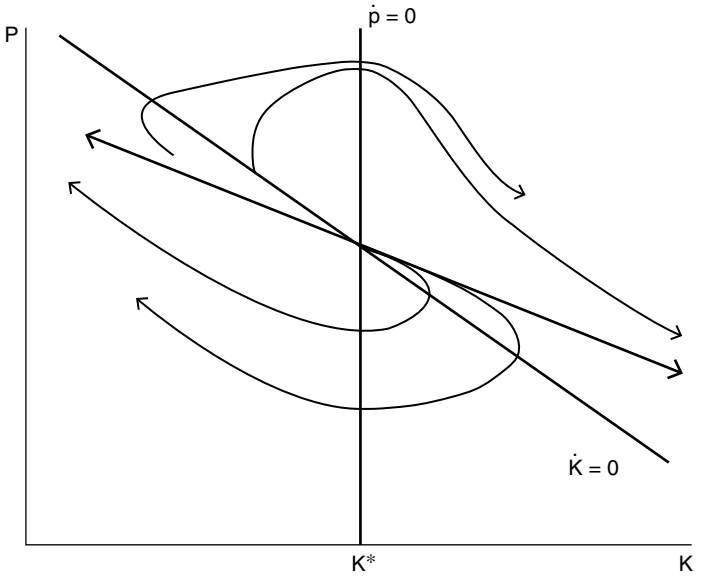


Figure 1.4a Unstable dynamics with non-productive land and non-linear savings function

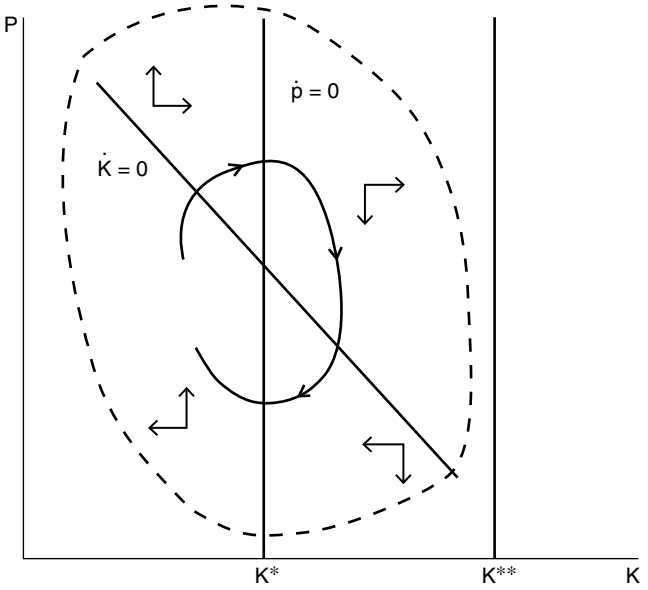


Figure 1.4b Oscillatory dynamics with non-productive land and non-linear savings function

if the $\left.\frac{dp}{dK}\right|_{\frac{dK}{dt}=0}$ locus hits the vertical axis, then the dynamics are constrained. $0 \leq K \leq K^{**}$ where K^{**} is defined by $f(K^{**})=\mu K^{**}$ (i.e. the capital stock that would result if the savings rate were unity.) p is non-negative. We can trace out a single oscillation along the path that begins say at $K=K^*$ and p very small. Such a path cannot hit the K^{**} boundary or the horizontal axis. If the value of p when it returns to K^* is lower than the initial p, then subsequent oscillations are arbitrarily close to the initial oscillation. If the value of p when it returns to K^* is greater than the initial p, all paths must be contained within the bound defined by this oscillation, a straightforward implication of which is that there must be a limit cycle.[100]

The second possible steady state is defined by $p=0$ and $s(K^{***},0,0)=0$. (Recall that $\frac{dp}{dt} = p(F_K - \mu)$ so that $\frac{dp}{dt} = 0$ for $p=0$ for all finite values of $F_K-\mu$.) If $s_k>0$, so long as p is constrained to be zero, the dynamics are stable. But if p is ever perturbed above zero, the dynamics described earlier become applicable.

1.7.2 Land-augmenting technological change

In this section, we assume that land is productive and the *effective* land supply increases at the rate n. The equation describing the equalization of returns to land and capital now takes on the form

$$\frac{dp}{dt} = pF_K - F_T \tag{7.9}$$

In steady state, $\frac{d}{dt}(\log p) = n$.

Because the rate of land-augmenting technical progress is n, one unit of land becomes more valuable over time at the rate n. We define

$$q=e^{-nt}p \tag{7.10}$$

so that

$$\frac{d}{dt}(\log q) = \frac{d}{dt}(\log p) - n = F_K - \frac{F_T e^{-nt}}{q} - n \tag{7.11}$$

Redefining units so that $T^\wedge$ is a unit of effective land, and denoting (as before) as output per unit effective labor, $f_{T^\wedge}=F_{T^\wedge}=F_T e^{-nt}$. Then the capital arbitrage equation can be rewritten

$$\frac{d}{dt}(\log q) = \frac{d}{dt}(\log p) - n = f_k - \frac{f_{T^\wedge}}{q} - n \tag{7.12}$$

In steady state, $\frac{d}{dt}(\log q) = 0$, so $f_k - n = \frac{f_{T^\wedge}}{q}$,

Or

$$q = \frac{f_{T^\wedge}}{f_k - n} \tag{7.13}$$

To simplify our analysis, for the remainder of this section we assume $\mu=0$ and we assume that a constant fraction of all income (including capital gains) is saved. We can write (7.2) as

$$\frac{d}{dt}(\log K) + \frac{pT}{K}\frac{d}{dt}(\log p) = s\left(\frac{F(K,L,T)}{K} + \frac{T}{K}\frac{dp}{dt}\right) \tag{7.14}$$

or in our normalized units

$$\frac{d}{dt}(\log k) = s\frac{f(k)}{k} - \frac{(1-s)(qf_k - f_{T^\wedge})}{k} - n \tag{7.14'}$$

The steady state is given by the solution to the loci along which $dq/dt=0$ and $dk/dt=0$, given respectively by[101]

$$q = \frac{f_{T^\wedge}}{f_k - n} = \Phi(\mathrm{k}), k < k^{**} \equiv f_k^{-1}(n) \tag{7.15}$$

and

$$q = \frac{sf(k) + (1-s)f_{T^\wedge} - nk}{(1-s)f_k} \equiv \psi(\mathrm{k}) \tag{7.16}$$

$\Phi'>0$ provided only that

$$\frac{F_{TK}}{F_T} > \frac{F_{KK}}{F_K}.$$

Under natural restrictions, the limit of Φ as k goes to zero is zero, and as k goes to k^{**} is infinity. In Figure 1.5, we have drawn the curve as upward sloping.[102] Above the curve, q is increasing; below it is decreasing.

$\psi(0)=0$ under natural restrictions. Again, under natural restrictions, for large enough k, the numerator of (7.16) becomes negative. Define k^{**} as the solution to[103]

$$sf(k^{**}) + (1-s)f_{T^\wedge} = nk^{**} \tag{7.17}$$

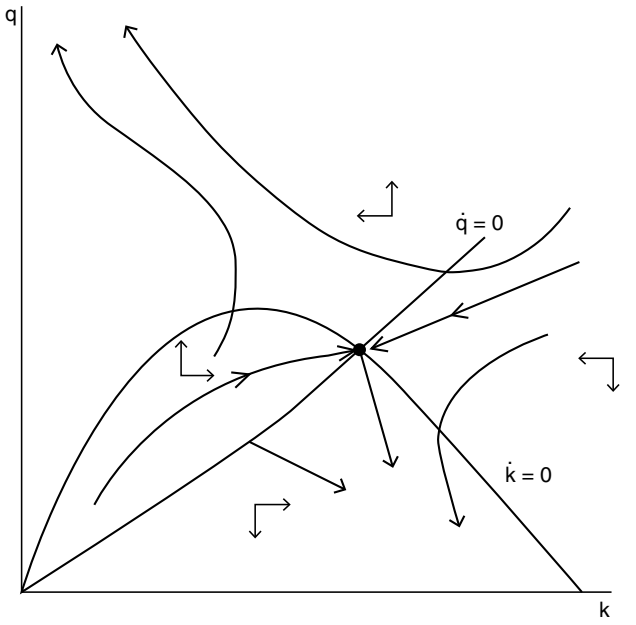


Figure 1.5 Steady state and unstable dynamics with land-augmenting technical change

Hence ψ is the inverted U shaped locus depicted in Figure 1.5. It is easy to show that the Φ locus cuts the ψ locus from below and there is a unique equilibrium. Above the locus, dk/dt is negative, below it is positive.

If land prices are too high, for ownership of land to generate the same returns as capital, the price of land has to increase. On the other hand, if q is above both the $dk/dt=0$ locus and the $dp/dt=0$ locus, it means that the increase in the value of land ("savings" in this sense) acts as a substitute for real capital accumulation, and k accordingly diminishes. *The result is that the steady state equilibrium is a saddle point, as depicted in the figure.*

With futures markets extending infinitely far into the future, q is set along the trajectory converging to the steady state, that is, there is a unique value of q for each k such that the economy converges to the steady state.

Without futures markets extending infinitely far into the future or infinite foresight, there is no reason to believe that the transversality condition will be satisfied. But along the paths which satisfy the short run arbitrage equation but do not converge to the long run equilibrium because the initial price is too high, the price of land eventually increases super exponentially.[104] As a result, in finite time, the "bubble" will be "corrected." But it can be a long time. And even when there is a "correction," it may still be on a "bubble path." The price of land falls, but to a level still above the convergent path.

Note that on the trajectories in which the price of land "explodes," eventually the increase in the value of land crowds out capital accumulation – the capital stock declines, *even though wealth continues to increase*. Indeed, as k gets small, virtually all of wealth is in the value of land, and thus wealth increases at the rate of $n+\frac{d}{dt}(\log q)=n+f_k-\frac{f_{T^\wedge}}{q}$. Above the $dq/dt=0$ locus, $qf_k>f_{T^\wedge}$ so that wealth is unambiguously increasing (and even increasing per capita). Indeed the wealth–income ratio (as usually defined, where income ignores capital gains) goes off to infinity.

Taxation. We now ask, what happens when we impose taxation on capital gains and/or the returns to land. The capital arbitrage equation now becomes

$$(1-t^{cg})\frac{dp}{dt}=pF_K-F_T(1-t^L) \tag{7.9'}$$

In steady state, the price of land is going up at the rate n, so in the steady state (using our normalized units)

$$(1-t^{cg})nq=qf_k-f_{T^\wedge}(1-t^L)$$

Or

$$q=\frac{f_{T^\wedge}(1-t^L)}{f_k-(1-t^{cg})n} \tag{7.13'}$$

To complete the analysis, we need to specify what is done with the revenues raised by the tax. Assume that they are entirely spent on consumption. Then the capital accumulation equation becomes

$$\frac{d}{dt}(\log k)=s\frac{f(k)}{k}-\frac{(1-s)(1-t^{cg})(qf_k-(1-t^L)f_{T^\wedge})}{k}-n, \tag{7.14''}$$

so in steady state

$$q=\frac{sf(k)-nk+(1-s)(1-t^L)(1-t^{cg})f_{T^\wedge})}{f_k(1-s)n(1-t^{cg})} \tag{7.16'}$$

The steady state is given by the solution to (7.13′) and (7.16′), giving the locus of $dq/dt=0$ and $dk/dt=0$ with land and capital gains taxes. From (7.13′) the land tax lowers the $dq/dt=0$ locus, but leaves the $dk/dt=0$ locus unchanged. As Figure 1.6a shows, this means that an increase in a tax on the return to land leads to an increase in the capital–labor ratio and an increase in wages,

validating the common presumption that savings diverted into land investment (or speculation) is money that could otherwise have gone into real investment.

A tax on capital gains shifts *both* curves upwards, and as Figure 1.6b shows, the consequence is again that the equilibrium capital–labor ratio increases. (The effect on the price of land is more ambiguous in the case of a tax on capital gains; along the $dk/dt=0$ curve, at any k, a higher tax on capital gains has to be offset by a high price of land, and by itself this would have implied a higher equilibrium q. But at the same time, this is partially offset by the shift downward of the $dq/dt=0$ locus.)[105]

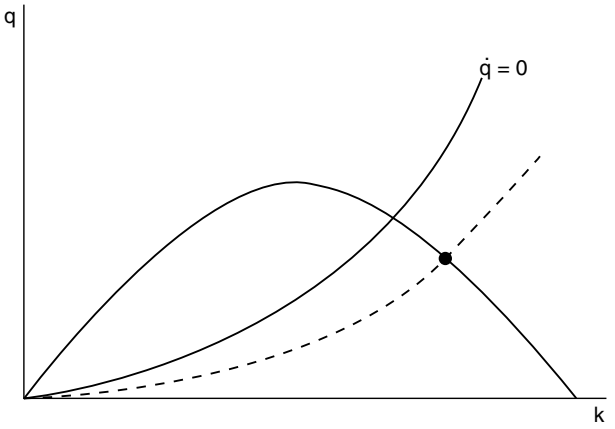


Figure 1.6a Effect of increase in tax on returns to land

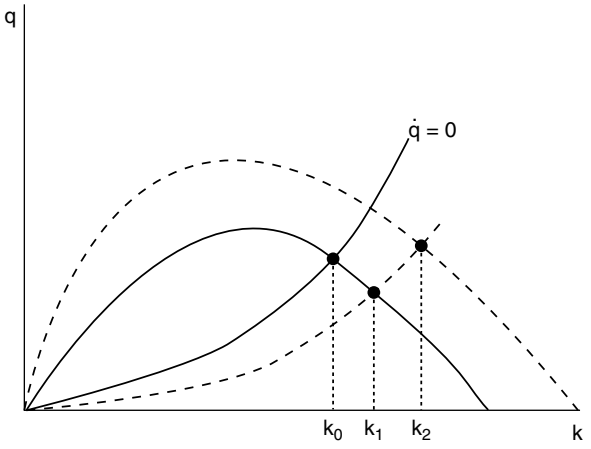


Figure 1.6b Effect of tax on capital gains

On the other hand, if the revenues are spent on investment, then

$$\frac{dk}{dt} = sf(k) - (1-s)(1-t^{cg})\ qn + t^{cg}qn + t^{L}f_{T^{\wedge}} - nk,$$

or in steady state

$$q = \frac{sf(k) - nk + t^{L}f_{T^{\wedge}}}{n(1-s)(1-t^{cg}) - nt^{cg}} \tag{7.16''}$$

Relative to the previous equilibrium, the $dk/dt=0$ curve is shifted up even more (while the $dq/dt=0$ curve is unchanged), so that the equilibrium value of k is increased even more.

1.8 Credit, collateral, and wealth inequality

We have argued that much of the growth of wealth is associated with the increased value of land and other fixed assets. Land, and certain other assets, have one attribute which makes them particularly attractive: they can be used as collateral. The fact that they can be used as collateral may increase their value; but the value associated with the ability to be used as collateral will depend on the financial system. If, for instance, banks do no lending based on collateral, then the "collateral value" will be zero; the easier access to credit for those who have collateral, the higher these assets will be valued.

Moreover, the demand for land and other assets depends itself on the availability of credit – a fact that was made abundantly clear by recent experiences with Quantitative Easing. (Indeed, one of the rationales for quantitative easing, and one of the main channels for its claimed success, was that it led to higher asset prices, with the hope that the increased wealth would in turn lead to more consumption.)

In this section, we suggest that the system by which credit is provided may be one of the main sources of wealth and income inequality: if a favored few get access to credit, then their wealth increases relative to those without such access. Nowhere was this clearer than in the former Soviet Union, where bank licenses were granted to some politically connected individuals. The access to funds that this provided enabled them to acquire state assets as they were being privatized; the limited access to funds meant that competition was limited and they could acquire the state assets at far below fair market value.

In a less dramatic way, wealth inequality in the United States and other advanced countries may also be linked with the financial system. If much of the growth of wealth is related to an increase in credit (or other changes

in the financial system); if access to credit is based on collateral; and if the assets which have benefited from the increase in credit (or other changes in the financial system) are disproportionately owned by the rich, then it should be apparent that these increases in credit and other changes in the financial system may have played a major role in the increase in wealth and income inequality.

Our system of credit creation may perversely create not only inequality at the top, but also at the bottom. It persuades the poor to borrow beyond their ability, and then charges them usurious interest rates. Changes in bankruptcy laws making it ever harder to discharge debts create a system of partial indebted servitude. Struggling to survive, they have no ability to make investments that would help them emerge from such poverty, and indeed, even to think long term. In the models below, we ignore these effects, focusing on the link between credit, collateral, land, and capital accumulation.

1.8.1 Credit and the value of land as a positional good

In this section, we provide a bare-bones model that we think may capture more accurately what has been going on than any of the models presented so far: the banking system provides credit based on collateral. When the price of land in the Riviera goes up, the banks are willing to lend more. If the banks are willing to lend more, the price of land in the Riviera goes up. There is, essentially, an indeterminacy: it is the decision of the banks (the central bank) concerning credit availability that drives the price of land (real estate).

We modify the model of section 2 by assuming three distinct classes of individuals – workers who just consume, capitalists who save out of profits, own enterprises and invest only in capital goods, but have no access to credit, and rentiers, who own land.[106] Their demand for positional good (land in the Riviera) is given by $M(W^T, \boldsymbol{c}, p)$, with the equilibrium condition now being given by

$$M(W^T, \boldsymbol{c}, p) = pT = W^T + \boldsymbol{c}, \tag{8.1}$$

where $\boldsymbol{c}$ is the amount of credit that is available and W^T is the wealth of the rentier, which is just the value of the land minus what they owe in credit: $W^T = pT - \boldsymbol{c}$. Substituting into (8.1) we obtain

$$M\left(W^T, \boldsymbol{c}, \frac{W^T + c}{T}\right) = W^T + \boldsymbol{c} \tag{8.1'}$$

We can solve for

$$p = \psi(\boldsymbol{c}) \tag{8.2}$$

The wealth of the rentiers is entirely driven by the provision of credit

$$W^T = pT - \boldsymbol{c} = T\psi(\boldsymbol{c}) - \boldsymbol{c} \tag{8.3}$$

To close the model, we need an additional equation describing capital accumulation. We take the simplest version, due to Kaldor (1957).[107] Capitalists-entrepreneurs save a fraction of their income, s_p, putting their money into capital goods

$$\frac{dK}{dt} = s_p rK - \mu K, \tag{8.4}$$

where μ is the depreciation rate, so in steady state

$$F_K(K^*, L) = \frac{\mu}{s_p}. \tag{8.5}$$

In this model, the provision of additional credit has *no* effect on the equilibrium capital stock. We thus obtain from (8.1′), letting $W = W^T + K$, the sum of the wealth of the rentiers and the capitalists,

$$\frac{dW}{d\boldsymbol{c}} = \frac{dW^T}{d\boldsymbol{c}} = \frac{-(1 - M_c - M_p/T)}{1 - M_{W^T} - M_p/T} \tag{8.6}$$

An increase in credit increases wealth through an increase in land prices, but has no effect on the capital stock. Since it is only the wealthy who own the land and that get access to credit, all of the increase in wealth (capital gain) goes to the wealthy. Monetary policy *causes* both the increase in (non-productive) wealth and the increase in wealth inequality. But while wealth has increased, wages are unaffected. Note that in this model, since credit simply leads to asset price increases (and an increase in the price only of the fixed asset land) – but not commodity price increases – there is no reason that a monetary authority focusing on commodity price inflation would circumscribe credit creation.

1.8.2 Credit and the creation of land bubbles and inequality

In this section, we consider a simple extension of the model of section 3 to incorporate credit, with land-augmenting technological progress at the rate n. To simplify, we assume that land and capital goods are perfect substitutes for each other, that there is no consumption value to land, and there are not two separate classes of entrepreneurs and rentiers. Land and capital are simply

alternative stores of value, and in equilibrium they must yield the same return. Then, as before,

$$\frac{d}{dt}(\log q) + \frac{f_{T^\wedge}}{q} = f_k - n. \tag{8.7}$$

Moreover, the full income of capitalists is now $F_K(pT+K)$, so that capital accumulation is described by (as before, letting $T^\wedge$ denote the effective land per worker, which is fixed, and assuming for simplicity that $\mu=0$)

$$\frac{dk}{dt} + T^\wedge(qf_k - f_{T^\wedge}) = s_p(f_k(T^\wedge + k) + T^\wedge(qf_k - f_{T^\wedge})) - n. \tag{8.8}$$

As before, (8.7) and (8.8) describe the full dynamics of the economy in terms of $\{q, k\}$.

Now assume, however, that the banking system[108] only provides credit with land as collateral, but provides it at zero interest rate, so that owners of land borrow as much as they can. The central bank limits the amount of credit that is made available. As more credit is provided, the price of land will be bid up, and in equilibrium

$$\boldsymbol{c} = \alpha pT. \tag{8.9}$$

where α reflects the collateral requirement. If α is fixed,

$$\frac{d}{dt}(\log p) = \frac{d}{dt}(\log \boldsymbol{c}) \tag{8.10}$$

Or

$$\frac{d}{dt}(\log q) = \frac{d}{dt}(\log \boldsymbol{c}) - n. \tag{8.11}$$

There is a path of expansion of the credit supply which ensures that (8.8) is satisfied. If the financial system expands credit supply at a pace that is faster than that implied by (8.8) and (8.10), the return to land will exceed the return to capital. In this polar model, if this were anticipated, no one would want to hold capital. The price of capital goods would fall below 1, and the production of capital would halt. *k* would decrease with the increase in the population. We then replace 8.7 with

$$\frac{d}{dt}(\log q) + \frac{f_{T^\wedge}}{q} = \frac{f_k}{z} + \text{ dlog z/dt} \tag{8.12}$$

where z is the price of capital goods in terms of consumption goods; and

$$\frac{d}{dt}(\log k) = -n. \tag{8.13}$$

k decreases and q increases.[109] If $\boldsymbol{c}$ increases fast enough, the value of wealth increases, and even wealth per capita increases, even though the capital stock per capita is decreasing.

Note that along such a trajectory the ratio of the (full) income of capitalists to that of workers will be increasing, provided that the elasticity of substitution is not too low (with the critical value being greater than one).

$$\frac{Y_K}{Y_L} = \frac{F_K(pT + K)}{F_L L}. \tag{8.14}$$

where Y_K is the (full) income of capital and Y_L that of labor. Note too that while the value of wealth is increasing, the return to capital will be increasing and that to labor decreasing. Hence trajectories where there is a rapid expansion of credit shift the income distribution towards capitalists. Of course, on such trajectories, growth in *output* will be low, in spite of the rapid increase in wealth. This simple model is consistent with the stylized facts described in Part I of this paper. (Note that while the ratio of full income of capitalists to that of workers is increasing, the ratio of income in the national income accounts to that of workers will be decreasing if the elasticity of substitution is less than one.)[110]

In more general models, where there is not a linear production possibilities frontier, an increase in credit leading to an increase in the value of land can initially lead to more investment, but eventually an increasing proportion of savings is absorbed by increases in the value of land, and, as here – and evidently as in many countries – the rate of *real* capital accumulation diminishes.

1.8.3 Credit creation, monetary policy, and inequality

To see more precisely how the "rules of the game" on credit creation can affect the distribution of wealth, first consider the model of the previous subsection, where credit is provided at a low rate against land as collateral. The return to holding land ρ_T is then the capital gain on land, the yield on land, and the option that owning land provides to get access to capital at a low rate:[111]

$$\rho_T = (1+\alpha)F_K = (1+\alpha)\left(\frac{d}{dt}(\log p) + \frac{F_T}{p}\right) \tag{8.15}$$

where here owning a dollar's worth of land allows one to borrow enough to increase one's land holdings to $(1+\alpha)$, on each unit of which one obtains

a return equal to the return on capital. In equilibrium, the return to land must equal the return to capital, and this means that if there is a change in the rules of the game – say a lowering of the collateral required for a loan – then there will be an increase in the price of land: those who are lucky enough to own land at that moment receive a large capital gain.[112] Such a change could be motivated by an improvement in the ability to manage risk, or by political influence, with the financial industry persuading politicians that such a change would allow a more efficient capital market. Of course, such a change in the regulations regarding lending does not increase the amount of real resources available in the economy, even if it might allow banks to lend more, and thereby might increase the profitability of banking.[113]

A slight variation of the life cycle model of Part III allows us to explore in more detail some of the distributive consequences of such a change or similarly, of a change in monetary policy that resulted in lower lending rates. Here, we investigate these issues in a highly stylized model that provides insights into the natural reasons that the ownership of land or other assets that might be used for collateral should be concentrated at the top. The issues can be seen more clearly in the context of a model where we assume only two factors of production, capital and labor, and that the ownership of capital ("equity in capital") can be used for collateral.

Assume that workers are very risk averse, while the wealthier capitalists are (close to) risk neutral. We assume that the government issues a fixed number of bonds B; each bond pays a fixed (real) interest rate, r_g, which is controlled by the government (monetary authority). We assume that the returns to capital are variable, so that all the capital is owned by the capitalists (they are the owners of equity), and all government bonds are owned by workers. Again, for simplicity, we assume that capitalists save and reinvest *all* of their gross income.The price of the bond is π. Thus the real rate of return to holding a bond is $\frac{r_g}{\pi}$. Because of risk aversion, $\frac{r_g}{\pi}$ can be substantially below $E(F_K)$, the expected return on capital, and workers will still hold their wealth in government bonds. On the other hand, so long as r_g/π is less than $E(F_K)$ no capitalist will hold a government bond. The price of the bond adjusts so that all of workers' savings is held in bonds,[114] i.e. assuming a constant savings rate of s out of wages (net of taxes). If workers pay no taxes, then

$$B\pi = sw \tag{8.16}$$

Interest on government bonds is financed through taxation. Not surprisingly, the structure of taxation matters.

Assume for simplicity that interest payments to workers are financed through a lump-sum tax τ on workers, i.e. $r_g\ B = \tau$. It can be shown that

equilibrium requires $\pi=1$,[115] i.e. $B=\frac{sw}{1+sr_g}$. Now, a change in r_g financed by a tax on labor leaves the returns to capital unchanged,[116] and that means that K^* is unchanged and w is unchanged; but it necessitates a change in B and τ. In particular, it can be shown that an increase in r_g leads to an increase in τ.[117] It thus leads to decreased first period consumption, but to increased second period consumption.[118] Since across steady states, $C_1+C_2=w$, the steady-state utility of workers is maximized at $r_g=0$ (when in effect individuals face the same constraint).[119]

In this model, the T bill rate is totally divorced from the rate of return on capital. We can, however, link the two, by assuming that the government, while borrowing from workers (who are engaged in life-cycle savings), is willing to lend to capitalists at a rate that is equal to or greater than that rate. For simplicity, we assume that there is a single rate, but that the government rations the amount it is willing to lend to capitalists, since so long as $r_g<F_K$, risk-neutral capitalists will want to borrow as much as possible. The way it rations credit is to require collateral. Hence, if a unit of capital allows a firm to borrow α, the overall return to a dollar of accumulation is $F_K(1+\alpha)-\alpha r_g$.

In the short run, a lowering of r_g leads to an increase in the net income of capitalists by an amount proportional to αK^* and a reduction of the income of life-cycle savers/workers by a corresponding amount. It is, in effect, a direct transfer from workers (life cycle savers) to capitalists.

Note that in this model, the distributive consequences of a lowering of the interest rate are the opposite of that derived in conventional "class" analysis, where workers are seen as debtors and capitalists as creditors. In that model, a lowering of the interest rate hurts capitalists and helps workers. Here, workers and capitalists are both owners of capital, but different kinds of capital. A lowering of the interest rate helps owners of equity and hurts those who hold government debt. This model seems to be a better description of the modern economy, and in this model, lowering interest rates unambiguously contributes to growing inequality. (This model, however, abstracts from Keynesian aggregate demand effects, which are the central motivation in lowering interest rates. We have assumed a full employment neoclassical economy.)

Over the long run, with α fixed, a lowering of r_g increases the return to investing, implying a higher equilibrium value of K,[120] and a higher wage rate, from which workers gain. The long-run equilibrium condition is (continuing with our simplifying assumption of $s_p=1$)

$$F_K(1+\alpha)-\alpha r_g=n \tag{8.17}$$

Moreover, as r_g is lowered, they gain also from the lowering of τ. But once r_g is lowered below zero, there is an offsetting distortion in the intertemporal

pattern of consumption. This means that there is (from workers' long run welfare perspective) an optimal $r_g < 0$.[121]

Inequality in wealth is given by sw/K; and it is possible to describe how this changes with a change in r_g.[122] For very large elasticities of substitution, the increase in K has little effect on w, so inequality increases; while for small elasticities of demand, the increase in K increases wages significantly, and reduces inequality.

Who gets the rents associated with credit creation? The essential insight of this analysis is that differences between life cycle savers and capitalists affect the asset composition of their holdings, and this means that policy changes (tax, monetary, and regulatory policies) affecting the relative returns and prices of different assets have differential effects on the two groups.

A natural question is, can't the process of credit allocation be changed to ensure that the rents associated with access to credit that are effectively being given the owners of capital through credit creation are more fairly shared? Why not have an auction of credit, so there won't be any rents?

Part of the answer is provided by the theory of information asymmetries: Stiglitz-Weiss (1981) and a large subsequent literature have explained why the provision of credit cannot be auctioned. There has to be an allocation process, entailing judgments about who is most likely to repay. But if that is the case, then who controls the allocation process makes a difference. Because it is a difficult task, entailing difficult judgments, it is natural that it be entrusted to those who are better educated, to the elites. But the elites are better judging those that are similar to themselves; there is an additional element of risk in judging those that are different. Moreover, there are shared judgments about risks and values. Not surprisingly, then, they allocate capital to those that are similar to themselves – even when and where connected lending is prohibited; and, of course, even more so when connected lending is allowed. In this manner, inequality builds on itself.

But that doesn't mean that there aren't *excessive* rents built into the financial system, and not just through the abuses that have been especially well-documented in the aftermath of the 2008 crisis, based on market exploitation (see, for example, Stiglitz, 2010). Consider, for instance, the allocation of credit for mortgages. Today, such allocation is not based on judgment so much as credit scoring. It is an information intensive process, involving the processing of information about the incomes of the borrower and the values of the properties being acquired. But government entities have the best data, and the government is in the best position to enforce the debt contract: the government, through the income tax system, has a complete history of income, and through property registries, of real estate transaction prices. The incremental cost of collecting mortgage payments through the income tax system is negligible. Indeed, it could easily construct a system of income contingent mortgage loans that

would be far better than the current system.[123] Administrative costs for such a system are likely to be very low, so that mortgages could be provided at an interest rate only slightly greater than that paid on government debt. The huge rents (and the associated instability and inequality) of the private mortgage system could be greatly reduced, and the enormous waste of resources as financial institutions look for fools upon whom they can prey would also be reduced.

1.9 Concluding remarks

This paper has attempted to explain key stylized facts about changes in macroeconomic variables, including those describing changes in the distribution of income and wealth, within models in which there is micro- and macro-consistency. Among the key findings are:

- Standard data on savings cannot be reconciled with the increase in the wealth–income ratio: there is a *wealth residual*. That is, observed increases in wealth and wealth–income ratios *cannot* be explained by the steady process of the accumulation of capital.
- An important component of the "wealth residual" is associated with an increase in rents: land rents, exploitation rents – including those associated with information asymmetries, monopoly and other forms of market power – and returns on intellectual property. Any theory attempting to explain the evolution of the economy must thus focus on explaining the increase in rents and their capitalized value, which are an increasingly important share of overall *wealth*.
- Concepts of "capital" and "wealth" are distinct. Appropriately defined aggregates for the inputs into production and wealth may be move in opposite directions. *The productive capacity of the economy may thus not increase in tandem with measured wealth.* Indeed, in many economies (including the US), productive capacity may be falling even as wealth is increasing. An increase in the market value of land does not make the economy more productive. An increase in wealth as a result of increased monopoly power, or political power which transfers resources from the public to corporations (for example, through corporate welfare) may even reduce the productive capacities of the economy.
- This resolves some long-standing paradoxes: the fundamental law of diminishing returns says that an increase in wealth should lead to a lowering of the return to capital. But, in fact, because *real wealth* – what might be called "real capital" or "productive capital" – has not increased in tandem with *measured wealth,* there is no necessity that the return to capital would fall as measured wealth increases.
- Similarly, we would expect an increase in wealth to be associated with an increase in average wages. This would be true even with technological

change (even if it is "skill-biased") or if there is a change in the composition of the labor force. The large gap that has opened up between the growth in appropriately measured average wages and productivity is consistent with the hypothesis that there has been an increase in market power.
- The increase in market power and other forms of exploitation rents can be in part explained by changes in technology and changes in institutional, legal, and economic structures.
- The paper provides several models (including life cycle models where land is a store of value or is a positional good) to explain why land values typically increase as wealth and wealth inequality increase: the wedge between wealth and real capital increases.
- The economy is dynamically unstable; that is, it is prone to the kinds of bubbles that have marked the economy in recent years. Particularly on such bubble paths, the increase in land values (wealth held in the form of land) crowds out real productive investment. Thus, the poor performance of the American economy – when employment, the growth of output, wages, median wealth and income, and other indicators are considered – should not come as a surprise.

In short, *this paper gives the theoretical underpinnings explaining why, in recent decades, wages have stagnated while GDP and productivity have continued to grow. It explains not only wage stagnation, but also credit-fueled bubbles. There has not been growth in productive capital relative to GDP (especially when controlling for increases in the value of real estate). The growth in wealth is for the most part simply the growth in the capitalized value of rents. This growth in rents has a negative effect on societal wellbeing.*

The paper also has explored key determinants in wealth distribution, focusing in particular on the distribution between life cycle savers and "capitalists." A marked change in the structure of the economy over the last 75 years has been the increase in life-cycle savings. We derive a simple formula describing the relative share of inherited wealth:

$$\frac{k^{w*}}{k^*} = n\frac{s(k^*)}{s_p}\frac{1-S_k}{S_k}$$

(where n is the rate of growth, s(k*) is the savings rate of workers, and s_p that of capitalists, and S_k the share of capital): In general, the wealth and income of capitalists will not continually increase relative to the rest of the economy, though in moving from one equilibrium to another, there can be marked increases. The magnitude of wealth inequality does not in general depend on the difference between the rate of return on capital or the rate of interest (r) and the rate of growth. The rate of return on capital is an endogenous variable, and needs to be related to underlying parameters of behavior and

technology (which the above formula does.) The analysis also notes that workers' savings does not (in this model) lead to higher wages or incomes per capita, but simply crowds out capitalists' savings. Introducing land into the standard neoclassical life cycle model lowers the equilibrium share of workers' wealth: wealth holdings in the form of land crowd out, in effect, productive savings.

We have also noted the ability of the financial system to exploit life-cycle savers, by lowering the return they receive on their assets relative to those received by the wealthy; this naturally decreases the share of wealth owned by life-cycle savers.

We have suggested, moreover, that there may be no fundamental difference between capitalists and workers: they may have the same savings *function,* with differences in savings rate simply reflecting differences in their wealth. It is not because workers save little that they have little wealth to pass on to their children; it is because they have little inherited wealth that they have a low savings. History matters: there can exist different steady states, depending on initial conditions.

In related work (Stiglitz, 2015), we have analyzed the distribution of wealth *among* capitalists.[124] The wealth distribution is the result of a balancing out of centrifugal and centripetal forces, forces pulling the economy apart and bringing it together. Increases in inequality can be attributed to the strengthening of centrifugal forces and the weakening of centripetal forces. Examples of strengthened centrifugal forces include the lowering of taxes at the top, the spending of more on the education of the children of the rich than of the poor, and increased dispersion of the returns to capital. An example of a weakening of centripetal forces is the weakening of public education. We are able to derive a simple formula describing the level of inequality in wealth among the very rich (the Pareto coefficient). It is related to the share of capital, the level of taxation of capital, and the degree of progressivity (or regressivity) of capital taxation. Again, because the return to capital is *endogenous*, the degree of inequality in the long-run equilibrium is *not* related to the difference between the growth of the economy and the return to capital. Moreover, just as the increase in wealth cannot be explained within the standard economic model, neither can the increase in wealth inequality. Once again, it is changes in policies, not economics, that are driving what is happening.

Our analysis provides a number of insights into how policies can affect inequality:

- Policies that reduce rents and their capitalized value can reduce inequality and increase economic performance. Efficiency and equity imply that rents should be taxed at very high rates; and taxes on capital with appropriate loss offset provisions actually encourage risk taking.

- Capital taxation has to be carefully designed to avoid problems of shifting (where before-tax returns increase, so that after-tax returns are not lowered). Progressive capital taxation with proceeds at least partially spent on public investments can reduce inequality and increase national output.
- A land (real estate) tax and a tax on natural resources – a *generalized* Henry George tax – can succeed not only in reducing inequality but can also lead to higher national output. (Since investments in land *crowd out* real investment, the lowered value of land as a result of taxation *crowds in* real investment.)
- Given the large amount of life-cycle savings, the traditional division of society into the owners of capital and workers or creditors and debtors may no longer provide the most insights for understanding the impact of policies on distribution. The relevant division is between capitalists, who pass on their wealth from generation to generation, and workers, who save for their retirement, and between the owners of equity and the holders of debt instruments. Since the wealthy are disproportionately the holders of equity, a lowering of interest rates (as in quantitative easing) benefits them but hurts holders of government bonds, disproportionately life-cycle savers, and thus increases inequality.
- We show the links between the increases in land values, monetary policies, and the structure and conduct of the financial system, demonstrating how changes in the rules governing that sector and the conduct of monetary authorities may increase inequality. A lowering of collateral requirements or of banks' capital adequacy requirements does not result in an increase in the overall efficiency of the economy, but leads to more inequality.[125]

In short, a tax on rents can raise revenue, not only incentivizing more productive investment, but also ensuring that more of society's scarce savings go into such productive investments, thereby enhancing growth and reducing inequality.

The deficiencies of the neoclassical model in explaining inequality that we have noted should make us wary about using that model for policy purposes – either for addressing inequality or for broader issues of economic performance. That model cannot account well for changes in inequality; we cannot explain these changes solely in terms of changes in the underlying key parameters that have traditionally been the focus of attention, related to technology and behavior, such as savings rates, bequest behavior, and reproduction rates, and the differences among families with respect to these variables.

For more than two centuries, there has been an attempt to break away from a feudal system in which a child's position in society is preordained by that of his parent, and to move to a meritocratic system where it is determined by the child's own ability. In many respects we have succeeded, but perhaps not as much as we had hoped: the evidence is that even in a society like the United States avowedly

committed to meritocracy, inherited advantages play a key role, and more than a role than can be explained by the process of transmission of genes.

We should be concerned with wealth inequality, however it is generated, because societies in which there are large wealth (and income) inequalities function differently from more equalitarian societies. There are social and political consequences. It is worth noting that the attack on monopolies and trusts in the Progressive era was more motivated by concerns about their political and social consequences than the market distortions to which they gave rise.

Because so much of the increase in inequality in income and wealth is related to changes in policies, changes in those policies may be able to ameliorate this growing inequality. If we believe that there are large costs to our economy, our democracies, and our societies from this growing inequality, then at the very least, we should ask, are there changes in policy which will slow down this increase in inequality – and perhaps reverse it. An understanding of the forces that may be contributing to the growing inequality, such as that we have attempted to provide here, is a first step in constructing such a policy agenda.

In fact, a long list of policy changes – changes in legal frameworks, taxes, and expenditures – which would lead to less inequality in both the short run and the long which might do this, and simultaneously increase economic performance, has already been identified.[126] It is not the lack of knowledge that is preventing these actions from being undertaken. It is politics, a politics shaped by inequality of political power which follows from and can amplify inequalities in economic power.[127] The growing inequality in our society is thus a reflection as much of democracy in the twenty-first century as it is of capitalism in the twenty-first century.

The fact that inequality is not just, or perhaps even mostly, the result of inexorable economic forces but of policies should be a source of hope: for it holds out the possibility that alternative policies might change the directions in which advanced economies seem to be heading. And it makes all the more imperative the research agenda to which this paper hopefully has made a contribution, of trying to understand better the determinants of the equilibrium wealth and income distribution.

Appendix A: Proof that average wages should increase with capital deepening

Assume $Y=F(K, L_1, L_2, \ldots)$ is constant returns to scale. In the following discussion, we will simplify and assume only two types of labor. Constant returns to scale (CRTS) implies that

$$F_{L_1}L_1+F_{L_2}L_2+F_K K=F,$$

so

$$F_{L_1,K}L_1+F_{L_2,K}L_2+F_{KK}K=0,$$

Diminishing returns implies $F_{KK}<0$, which is why if there is only one type of labor $F_{LK}>0$: an increase in capital must increase the marginal productivity of labor, so that an increase in capital (relative to labor) must increase the wage. Here, it is clear that the wage of one of the two types of labor could go down.

But consider the average wage, $\bar{w}$:

$$\bar{w}(K)=(F_{L_1}L_1+F_{L_2}L_2)/L$$

where $L=L_1+L_2$.

$$\bar{w}'(K)=\frac{(F_{L_1,K}L_1+F_{L_2,K}L_2)}{L}=-\frac{F_{KK}K}{L}>0.$$

The weighted average wage must increase when capital (the capital labor ratio) is increased.

This result is strengthened if we assume that there is an increase in the quality of the labor force. Let ω be the proportion of high skilled workers.[128] Then

$$\frac{d\bar{w}}{dt}=\bar{w}_K\frac{dK}{dt}+(F_{L_2}-F_{L_1})\frac{d\omega}{dt}>0.$$

Appendix B: Relationship between change in wealth and aggregate inputs

Consider the rate of change of wealth, $W=K+pT$:

$$\frac{dlog(W)}{dt}=\xi\frac{dlog(K)}{dt}+(1-\xi)\frac{dlog(p)}{dt} \tag{B.1}$$

where it will be recalled

$$\xi=K/W. \tag{B.2}$$

But

$$p=[(1-\zeta)/\zeta]K/T, \tag{B.3}$$

so, substituting into (B.2), we obtain $\xi\equiv\zeta$ so that

$$\frac{d}{dt}(\log(\boldsymbol{C}))=s/\beta \tag{B.4}$$

and

$$\frac{d}{dt}\left(log\left(\frac{\boldsymbol{C}}{Y}\right)\right)=\frac{s}{\beta}-g \tag{B.5}$$

Moreover,

$$\frac{dlog(p)}{dt} = \frac{dlog(K)}{dt}.$$

Hence

$$\frac{dlog(W)}{dt} = \frac{dlog(K)}{dt}, \tag{B.6}$$

Appendix C: Inheritance taxes

In this appendix, we analyze the effects of taxing only the return on inherited wealth. Life cycle savings is exempted, e.g. through IRA accounts. Now, we have a somewhat more complicated problem than that analyzed in the tax. Tax revenues are given by

$$T = \tau^c r(k^*)(k^* - k^{w*}) \tag{C.1}$$

where

$$k^{w*} = s(k^*)(w(k^*) + T). \tag{C.2}$$

Substituting (C.1) into (C.2), we obtain

$$k^{w*} = \frac{s(k^*)(w(k^*) + \tau^c f'(k^*)k^*)}{1 + s(k^*)\tau^c f'(k^*)} \tag{C.3}$$

We have already shown that as τ^c increases $w(k^*) + \tau^c f'(k^*)$ decreases. Similarly, as τ^c increases the denominator increases. Hence, so long as $s' \geq 0$, k^{w*} decreases. k^* decreases. If the elasticity of substitution is greater than a critical threshold (less than unity), the share of life-cycle wealth increases; but if the elasticity of substitution is very small, it can decrease because of tax shifting.[129]

Now, however, the effect on relative consumption (wellbeing) is more ambiguous. In particular, at $\tau^c = 0$, using (3.11)

$$\frac{dY^W}{d\tau^c} = \frac{dw}{dk^*}\frac{dk^*}{d\tau^c} + r(k^*)(k^* - k^{w*}) = -k^{w*} f'(k^*) < 0. \tag{C.4}$$

On the other hand, since $r(k^*) = f'(k^*) = \frac{1}{s_p(1-\tau^c)}$, $\frac{dr}{d\tau^c} = \frac{r}{1-\tau^c}$. Workers' lifetime utility if a function of their income and the return to capital: $V(r(k), Y^w)$, where V is the indirect utility function.[130] Hence[131] at $\tau^c = 0$,

$$\frac{dV}{d\tau^c} = \frac{\partial V}{\partial Y^w}[k^{w*} f'(k^*) + (-k^{w*} f'(k^*)] = 0. \tag{C.5}$$

That is, the loss in income is precisely offset by the increased return to capital. But for $\tau^c = 0$, the interest rate effect is larger, and initially the transfers are larger, and workers'

utility is increased, even though wages are lower. But as τ^c increases, eventually k^* falls below k^{*w}: the economy switches to a one class economy, with only life cycle savings, with

$$s(k^{w*})w(k^{w*}) = \frac{k^{w*}}{1+n}.$$

Clearly, because wages are lower than they were in the initial equilibrium and there are no transfers, workers' incomes are lower. There exists an optimal inheritance tax τ^{c*}, $0 < \tau^{c*} < 1$.[132]

Notes

1. University Professor, Columbia University. This is a revised version of a paper originally presented at an IEA/World Bank Roundtable on Shared Prosperity, Jordan, June 10–11, 2014 and at an INET seminar at Columbia University, December 3, 2014. I am grateful for the helpful comments of the participants in the roundtable and seminar, and in particular, to the discussants, Duncan Foley, Paul Krugman, and Banko Milanovich.

   The issues discussed in section 2 of this paper on the measurement of wealth and capital were discussed at a special session of the IEA World Congress, Amman, sponsored by the OECD on the Measurement of Wellbeing, and at a meeting sponsored by the OECD High Level Expert Group on the Measurement of Economic Performance and Social Progress, Rome, September 2014. I am indebted to Martin Durand, Chief Statistician of the OECD, and other participants at those meetings for their helpful comments and insights into the key issues of the measurement of wealth and capital, and in particular to Paul Schreyer, both for his insights and for supplying me with the data cited in section 2. I am also indebted to Martin Guzman, Arjun Jaradayev, Suresh Naidu, Stefano Battiston, and Mauro Gallegati for their comments. I have also benefited from conversations with Adair Turner and Shahe Emran. My earlier work in this area was greatly influenced by Tony Atkinson, David Bevan, John Flemming, Robert Solow, James Meade, Frank Hahn, Nicholas Kaldor, Jim Mirrlees, Benoit Mandelbrot, and David Champernowne. Financial support was provided by INET (the Institute for New Economic Thinking) and the Ford Foundation Inequality Project at Roosevelt Institute. I am indebted to Feiran Zhang, Ruoke Yang and Eamon Kircher-Allen for research assistance.
2. Kaldor (1961). For a recent review of the attempts to explain these facts, see Jones and Romer (2009).
3. This paper focuses on advanced countries. A slightly different analysis would be required for developing and emerging countries, though the theoretical formulations presented here are general.
4. These are not the only stylized facts that need to be explained. There is a large literature trying to explain the *shape* of the income and wealth distribution, for example, why the tails of the distribution are Pareto (fat-tailed), and why at lower levels of income, the income distribution seems to be described by a lognormal distribution (Aitchison and Brown, 1957; Lebergott, 1959). These are addressed in Stiglitz (2015b).
5. Reflected not just in growing Gini coefficients, but an increasing share of income going to those at the very top. Given that inequality in each of the components of

income (wages and the return to capital) is increasing, and the relative importance of the more unequal component, capital, is also increasing, it is, of course, obvious that there would be an increase in overall level of income inequality. See, for example, OECD (2011) and Piketty (2014).

6. Real US wages have stagnated for decades (see Shierholz and Mishel, 2013). Adjusted for inflation, average hourly earnings of production and non-supervisory employees have decreased some 30 percent since 1990. See St. Louis Fed data at http://research.stlouisfed.org/fred2/series/AHETPI/. More dramatic, while the labor share may have decreased from the mid 80 percent in the 1970s to less than 80 percent by 2009, the aggregate labor share excluding the top 1 percent compensation (whose returns, as we note below, often consists significantly of what can be referred to as rents) has slid from just under 80 percent to around 60 percent. See Giovannoni (2015).
7. See Piketty (2014) and Piketty and Zucman (2014). For the UK, Germany, and France, the wealth–income ratio rose from about 300–360 percent in 1970 to 377–618 percent in 2010. The US had a relatively small increase, from 399 percent to 456 percent.
8. Kuznets (1955).
9. Piketty also says that, among those who hold wealth, "the distribution of wealth tends toward a long-run equilibrium and that the equilibrium level of inequality is an increasing function of the gap $r-g$ between the rate of return on capital and the growth rate … The greater the difference $r-g$, the more powerful the divergent force. If the demographic and economic shocks take a multiplicative form (i.e., the greater the initial capital, the greater the effect of a good or bad investment), the long-run equilibrium distribution is a Pareto distribution (a mathematical form based on a power law, which corresponds fairly well to distributions observed in practice). One can also show fairly easily that the coefficient of the Pareto distribution (which measures the degree of inequality) is a steeply increasing function of the difference $r-g$." (Piketty, 2014, pp. 363–4). We examine these hypotheses in Stiglitz (2015b), showing that the qualitative propositions are not, in general, valid.
10. Roberts (2014). It appears that Piketty's analysis seems to have overestimated "r", overestimated the extent to which returns were reinvested, and underestimated the importance of the division of wealth among one's heirs.
11. Piketty himself recognizes the possibility that there can be an increase in the value of land, but quickly dismisses its historical importance (though he notes that does not mean that its importance might rise in the future): "… the increase in the value of pure land does not seem to explain much of the historical rebound of the capital/income ratio (sic)" (p. 198).
12. Included in the increase in the value of land is the value of artificially created scarcity, e.g. through zoning requirements. Land rents are likely to go up significantly with increasing urban agglomerations—it is not, as Piketty (2014) seems to suggest, that rents some places go up, and others go down. For instance, in a simple model of the city, Arnott and Stiglitz (1979) show that land rents go up with aggregate transport costs. Not surprisingly, the importance of agglomerations increases with the size of local public goods. (In their highly idealized model, they obtain the result that with cities of optimal size, differential land rents are equal to the expenditures on local public goods, and are one half the value of aggregate transport costs.)
13. A result that is consistent with the findings of Galbraith (2012).
14. See, for example, Stiglitz (1966, 1969).
15. At least for some countries, there appears to be an increase in inherited inequality relative to life cycle inequality Bowles and Gintis (2002) and Piketty (2014). But

there is not unanimity about this conclusion. See, in particular, for the US, Wolff and Gittleman (2011). Our model enables us to ascertain the conditions under which either result might be expected.

16. As we shall note below, what really matters is the growth of the effective labor force, the sum of the labor force growth rate and the rate of labor augmenting change.
17. It is worth noting that in standard models, the condition $r \geq g$ must be satisfied if the economy is intertemporally efficient. If Piketty's analysis were correct, it would imply that, except in the limiting case where r=g, *any efficient economy would be characterized by ever increasing inequality*.

    It is also worth noting that in the special parameterization so loved by macroeconomists, the Cobb–Douglas production function, average and marginal returns move in tandem, so that a fall in the average productivity of capital would be accompanied by an equiproportionate fall in the marginal productivity. In the case of an elasticity of substitution less than unity, the fall in the marginal productivity is larger. (See the discussion below.)
18. More precisely, as we will explain below, in the effective-capital labor ratio, taking into account the increased productivity of each worker.
19. There are still other anomalies about which we will have only a little to say in this paper. Globalization was supposed to increase societal welfare *for all countries*; even if there were distributional effects within countries, the gainers could more than compensate the losers. There is increasing evidence that there are indeed losers (Acemoglu et al., 2014); but the losers are being told that they must accept *further* cutbacks in wages and government services *in order for the country to compete*, seemingly suggesting that globalization requires them to accept a lower standard of living.
20. See Arrow et al. (1961); Young (2013). It should be noted that some authors have recently argued otherwise. See, for example, Mallick (2007).
21. The net private savings rate for the US over the period 1970–2010 has been 7.7 percent (Piketty and Zucman, 2014). As they point out, most of the variability in wealth income ratios (at least as conventionally measured) can be attributed to the private sector.
22. This can be expressed in another way. The average annual increase in the capital stock for the US they estimate to be 3.0 percent, of which the average "real" savings accounts (by the calculation above) to about 1.5 percent, or just half. (Piketty and Zucman (2014) suggest that savings accounts for 72 percent of the increase in the wealth–income ratio.)
23. We obtain similar results if we postulate particular behavioral models. Take a simplified version of the model that seems to underlay Piketty's analysis, a Kaldorian savings model (Kaldor, 1957), where capitalists save a fraction s_p of their income and workers nothing. Piketty (2014) *implicitly* seems to assume s_p=1, but the overwhelming evidence is that even the very rich save a much smaller fraction of their income than that. Saez and Zucman (2014) estimate that the average saving rate for the wealthiest 1 percent of Americans was 36 percent from 1986 to 2012. Similarly, Dynan et al. (2004) obtain high savings rates for the rich – but far lower than unity. For simplicity, assume the *after-tax* rate of return on capital (it should be obvious that what matters is after tax returns) is 5 percent, s_p=.4. Then capital would increase at the rate of .05×.4=.02. If the growth rate were greater than 2 percent, the private capital–output ratio would be declining. Note that if the share of capital is around .2, this generates a national savings rate of 8 percent, just slightly higher than the actual private savings rate.

Similar results hold if there are some savings out of wages. As Pasinetti (1962) notes, a more reasonable model divides income according to whom it accrues, i.e. interest and wage income accruing to workers is treated similarly. In Part III of this paper, we assume workers save for their retirement, while capitalists save to pass on money to their heirs. In Part III, we sketch a model in which the division of society into these different groups arises endogenously.

24. It is obvious, of course, that the short run fluctuations in the wealth–income ratio are dominated by capital gains and by cyclical movements in income. The marked changes in the wealth–income ratio in the US before and after 2008 highlight these points.
25. In the Kaldorian model, the long-run capital–output ratio is given by $s_p S_K / g^*$, where S_K is the share of capital.
26. For instance, between 1960 and 2000, the savings rate fell from 8 percent to 2 percent while the rate of growth increased from 2.3 percent to 4.1 percent. If these were permanent changes, then the long run capital–output ratio would have fallen by a factor of almost 8. (Actually observed growth rates will be higher than g^* – the sum of the rate of growth of population and labor augmenting technological progress – if there has been capital deepening, less than g^* if the reverse has been happening.
27. Matters are no better if we view the savings rate as endogenous, determined by intertemporal utility maximization. Then, the critical variable is the intertemporal discount rate, and again, it is hard to see changes in that variable of the magnitude that would account for changes in the observed capital–output ratio.
28. For wage data see Shierholz and Mishel (2013).
29. That is, Giovannoni (2014) noted that between 1980 and 1990 the share of the bottom 99 percent of workers has gone down by over 20 percent, which means that the ratio of their average wage to their average productivity has gone down by the same amount. More dramatic results are observed if we look at broad categories of workers like production and non-supervisory workers, where (real) wages have stagnated over the past forty years, while average productivity has doubled. Note that with the Cobb–Douglas production function much beloved by macro-economists, marginal and average productivities move perfectly together. Note too that skill-biased technological change might explain why there might be marked disparities in movements in median wages and average productivity; but it does not explain the phenomenon just described.
30. The first to propose the idea of skill-biased technological change was Griliches (1969). See also Krusell et al. (2000), Autor (2002); and Autor, Katz, and Kearney (2008).
31. See Card and DiNardo (2002) and Shierholz, Mishel, and Schmitt (2013) and the references cited there.
32. Interpretations of Piketty's work, which confuse the increase of "wealth" with an increase in capital argue that there *must* be an elasticity of substitution greater than unity – how else could one explain the rising share of capital. But if the elasticity of substitution is greater than unity, then labor augmenting technological change would lead to an increase in wages at a fixed capital stock, and an even larger increase in wages were the capital stock to increase. (The elasticity of substitution has to be substantially below unity for the wage to decrease. If there are different kinds of labor, similar results hold for the average wage.)
33. Labor augmenting technological change leads to a higher return to capital, and the presumption is that it would lead to higher investment. This would lead to a still higher wage.

34. For simplicity, assume that only the productivity of skilled workers (denoted with subscript 1) increased. Let λ=the productivity of a skilled worker. If $\frac{dF_K}{dt} = 0$, $F_{KK}\frac{dK}{dt} + F_{1,K}\lambda L_1 \frac{dlog(\lambda)}{dt} = 0$, and normalizing L at unity, $\frac{d\bar{w}}{dt} = (F_{1,K}\lambda L_1 + F_{2,K}L_2)\frac{dK}{dt}$ $+(F_{1,1}\lambda L_1 + F_{2,1}L_2)\lambda L_1 \frac{dlog(\lambda)}{dt} + F_1 L_1 \lambda \frac{dlog(\lambda)}{dt}$. Using the properties of constant returns to scale production functions and the condition that F_K is unchanged, we can show that $\frac{d\bar{w}}{dt} = F_1 L_1 \lambda \frac{dlog(\lambda)}{dt} > 0$.
35. The analysis of capital-augmenting technological progress is somewhat more complicated. First, the "volume" measure of the capital stock discussed below is supposed to adjust for differences in quality of capital. Whether it does so adequately is beyond the scope of this paper. Secondly, with capital augmenting technological progress, there is no steady state. Short-term capital augmenting progress, by increasing the effective capital stock, would have been expected to have an unambiguously positive effect on wages.
36. These problems are similar to those that have arisen in the measurement of poverty, with Pogge and Reddy (2010) arguing that standard estimates do not adequately reflect differences in prices faced by the poor – a claim that Martin Ravallon has disputed, illustrating that these index number problems are both difficult and contentious.

    See, in particular, the discussion of positional goods later in this paper.
37. I am very indebted to Paul Schreyer of the OECD, who concludes his discussion of these issues (personal note to author) by observing "the distinction between the wealth and production aspects of capital is indeed important and a story about '*W*' does not immediately translate into a story about '*K*'. Associated with the two perspectives are different measures that evolve quite differently. However, the key aspect in the analysis of capital in production and its link to income shares seems to be the treatment of non-produced assets, in particular land."
38. As we noted earlier, although land is not very important in most industrial processes (certainly not as important as it is in agriculture), housing services represent an important component of GDP, and land is an important input into real estate.
39. We note that this is not a plausible production function, since if that were the case, there shouldn't be any changes in the relative price of *T* and *K*, since they are perfect substitutes.
40. Similar results hold for the two other countries for which we have been able to obtain comparable data, Australia and Korea, from the OECD. Land accounts for a large part of national wealth – at current prices, between 40 and 60 percent – and the wealth–output ratio excluding land has been rising, while the ratio including land has been falling. I am indebted to Paul Schreyer for these data.
41. See Stiglitz (2003, 2010a) and the references cited there.
42. Indeed as Giovannoni (2014) points out, simply excluding the top 1 percent of wage earners results in a very large decline of the wage share between the period from around 1980 to 2009, from slightly more than 75 percent to around 60 percent.
43. Some question the magnitude of some of the increase in inequality, say the share of income at the top for the US, because of changes in the tax law in 1986 which may have led to a change in *reported* income, not actual incomes earned. (Feldstein, 2014) We should note that the studies of inequality looking at the increased inequality at the top have attempted to deal with this obvious problem. (Piketty and Saez, 2003).

But the pattern of increased inequality (an increased share of total income going to the top 1 percent) continued even after tax changes were partially reversed in 1993. Moreover, other countries without corresponding changes in tax codes have seen similar increases in inequality. (Interestingly, because in the US, the top is the only part of distribution that has done very well, if it were the case that most of their seeming increase in income is just a change in reporting, it would imply that that the overall performance of economy has been really dismal; one would have to explain how it is that, given all of the increase in wealth, all of the "improvements" in economic policy, and all of the alleged gains from globalization and technology, all of these together seem to have generated so little improvement in standards of living to any group in our society, not even, allegedly, the very top.)

It is, of course, plausible that the overall level of inequality at the top is *greater* than that reported. Administrative data show *reported* (realized) capital gains, but the tax system provides strong incentives for those at the top not to realize their capital gains.

44. In the short run, there can be capital gains on producible assets as well, but such increases cannot be sustained in the long run, since they will elicit a supply response. Some of the increase in "seeming" wealth that occurred in the US prior to the 2008 crisis may have been attributable to capital gains on buildings (though it is difficult to parse out such capital gains from capital gains on land). But the "correction" brought down the implied price of building to or below the reproduction cost. If we take consumption goods as our numeraire, the price of capital goods could increase or decrease, though such changes typically are of a limited magnitude in the absence of technological change; with technological change, there can, of course, be significant changes in appropriately measured prices.
45. Itself an endogenous variable. Changes in preferences and technology can lead to increased agglomerations, with an increase in land values.
46. See Knoll, Schularick, and Steger (2014, 2015).
47. Hotelling (1931) showed that if the cost of extraction of a depletable natural resource were zero, its price would rise at the rate of interest (which in an efficient equilibrium is always greater than or equal to the rate of growth.)
48. Piketty, Saez, and Stantcheva (2014) provide an interesting empirical test, pointing out that increases in tax rates at the very top are *not* associated with slower rates of growth. See Stiglitz (2012a, 2014b) for a broader discussion, including the many forms that rent-seeking takes in a modern economy, and other evidence that rents have become an important source of income at the very top.
49. The timing of increases in the share of capital are perhaps more consistent with those being explained by rapid changes in the degree of exploitation than by sudden changes in the effective capital labor ratio. Similarly, it is hard to reconcile the enormous divergence between average compensation and productivity of workers *without* assuming an increase in market power. (See Giovannoni, 2014.)
50. Assume, for instance, that $W/Y = 4$. Assume the increase in rents are capitalized in the stock market. Then $\Delta W = .05Y/.015$, so if Y is unchanged, $\Delta(W/Y) = 3.33$, so now $W/Y = 7.333$. Actually, the increase in the wealth–income ratio is even greater than these calculations would suggest, since, as we note in the next paragraph, the distortion in the economy lowers the magnitude of the denominator.
51. See, for instance, Federal Reserve Board (2015) for a discussion of the cost to consumers of predatory lending practices.
52. Indeed, the extensive research on efficient markets has questioned the value-added of the wealth management services of the financial sector: ordinary investors would have done as well or better simply by buying indexed funds.

53. Stiglitz (2012a) outlines many other forms of rent seeking. Some forms of rent-seeking may detract from measured wealth. If CEOs are able and willing to take greater advantage of deficiencies in corporate governance laws to appropriate for themselves more of the value of corporations, that *should* lead to a decrease in the market value of firms. There is, however, considerable evidence that because of the lack of transparency of the manner in which they appropriate these returns, markets typically do not fully reflect the dilution in shareholder value. Moreover, much of the compensation takes the form simply of a transfer of ownership claims on the returns to the firm. Note further that if this rent appropriation by managers is labeled as "compensation," then the wage share is increased. This is consistent with the results noted earlier suggesting a marked decline in the wage share if the upper one percent of "wage earners" are excluded.
54. This discussion raises similar issues as those the Commission on the Measurement of Economic Performance and Social Progress discussed in moving economic activities from the public to the private sector (see Stiglitz et al., 2010).
55. See, for example, Stigltiz (1975).
56. See, for instance, Henry and Stiglitz (2010) and the works cited there.
57. As we note below, such changes are often accompanied by a loss in wellbeing of others: they must now make royalty payments to the owner of this intellectual property. But the diminution of their wellbeing is not necessarily reflected symmetrically in the wealth accounts. Moreover, the charges imposed for the use of knowledge lower GDP, and thus a change in the intellectual property regime extending rights to enclose the knowledge commons can both increase the measured value of wealth and lower the value of GDP: the wealth–income ratio will accordingly rise.

    The privatization of public knowledge or the granting of "excessive" intellectual property (patents that are excessively broad, such as covering all four-wheeled self-propelled vehicles, or copyrights that are excessively long, such as extending 70 years beyond the death of the writer) can be viewed as a special case of the exploitation rents discussed above.

    There is one more form of rents associated with intellectual property that has almost surely grown over time: that generated by brand names, especially the identification of a product with say a sports star).
58. See Sraffa (1960) and Stiglitz (1974). Thus, in models with the production of commodities by means of commodities, the economy at a low interest rate and a high interest rate may look the same (the same technologies are employed), while at an intermediate interest rate a different technology is employed. Even if the value of wealth has changed in going from the low to the high interest rate, there has not been capital deepening, at least in any meaningful *real* sense. There are a variety of other reasons that there can be changes in intertemporal pricing, with large consequences to the valuation of assets. See the discussion below.
59. While financial markets often claim that their innovations have enhanced the ability to manage risk, the extent to which this is the case remains debated. Some of the financial innovations may have actually increased risk (Stiglitz, 2010b). Some of the financial innovations may have led to the creation of pseudo-wealth – wealth based simply on differences in perceptions in beliefs (Guzman and Stiglitz (2014)); while other innovations, like improvements in the ability to sell short, may reduce market values (Scheinkman and Xiong, 2003). Part IV of this paper will show how changes in financial market regulations can affect the value of assets.
60. See Milevsky and Huang (2011). For statistics on the size of pension funds, see OECD (2013).

61. In that sense, the model is similar to that of Pasinetti (1962), where there are two classes too. We model workers' saving (life cycle savings).
62. S_p can be derived endogenously, if, as in the standard representative agent model, families maximize dynastic utility.
63. Notice that for capitalists, savings are defined as the *addition* to their wealth, while for workers, since each worker starts life (in this model) with no wealth, savings are their total wealth. (There are alternative formulations based on gross savings generating similar results.)
64. We could have employed a more general savings function: $S(k_t, k_{t+1})$ where the savings rate depends not only on the rate of return on capital (which depends on k_{t+1}) but also on wages, which depend on k_t. It should be apparent that in the steady state, savings is just a function of k. Little here depends on the precise form of s, though we will observe that some results do depend on whether savings increase or decrease with k_t. Note that an increase in k will be associated with an increase in wages and a decrease in interest rates. s will increase with k so long as the substitution effect of the decreased wages is not too large.
65. As Stiglitz (2010b) shows, there can in general be an infinite number of trajectories consistent with rational expectations. This follows from the fact that there may be more than one solution to (3.1) and (3.2) and (3.3) for k^w_{t+1} for any k_t. (Substituting (3.1) into (3.2), we obtain $k^w_{t+1} = s(k^w_{t+1} + \theta(1 + s_p f'(k_t))k^c_t/(1+n))w(k_t)$). The reason is that if workers expect a high interest rate, they will need to save little for their retirement – but then the interest rate will be high; but if they expect a low interest rate, they will need to save a lot, but then the interest rate will be low.
66. If workers' intertemporal utility functions are Cobb-Douglas, then s'=0. If workers' utility function is such that U=min $\{C_t, C_{t+1}\}$, then (1–s)w=s(1+r)w, or s=1/[2+r], so (3.7) can be rewritten $\frac{k^{w*}}{k^*} = n\frac{1}{2s_p + n}\frac{1-S_k}{S_k}$. An increase in s_p reduces the share of inherited wealth provided the elasticity of substitution is not too small.
67. As we have noted earlier, there are a number of other factors that could affect life cycle savings – the adequacy of provision of health care for old age, the efficiency of annuity markets and the extent to which they are affected by asymmetries of information, and uncertainties both about retirement age, rates of return to capital, and life expectancies. In practice, there are other institutional factors: most individuals save through retirement programs, and the rules and regulations concerning those retirement programs can have first order effects on the amount set aside.
68. The critical condition is that $s(k^*)w(k^*) < k^*$, or that $\frac{s(k^*)}{s_p} < \frac{S_k}{n(1-S_k)}$. If $n=1$, $S_k=0.2$, then the condition becomes $s(k^*) < 0.25 s_p$.
69. We should emphasize that this result is not general. In Part IV of this paper, we consider, for instance, a model in which capitalists have a choice of assets to hold, and in equilibrium, they hold all of the risky assets. In a generalization of that model, it is easy to show that a tax on the excess returns to capital over the safe interest rate leads to more risk taking, i.e. a shift in their portfolio to higher return assets (Domar and Musgrave, 1944; Stiglitz, 1969b). If these assets are complements to labor, that shift by itself may increase wages. We note later too that taxes on capital gains in land may redirect investment into forms that are more complementary with labor.
70. From (3.4a) $\frac{f''(k)k}{f'(k)}\frac{dlog(k)}{dt} = \frac{\tau^c}{1-\tau^c}\frac{dlog(\tau^c)}{dt}$

71. Since s is fixed, and Y^w falls, k^{w*} falls, while k* increases. We can rewrite (3.7) with taxes as $\frac{k^{w*}}{k^*} = n\frac{s(k^*)}{s_p(1-\tau^c)}\left(\frac{1-S_k}{S_k}+\tau^c\right)$ where S_k is the share of capital before tax.
72. If the government invested only a fraction z of its revenues, then if z is small enough ($<s_p r(1-\tau^c \equiv z^*$), there is an equilibrium ratio of $\frac{K_p}{K_g}$ given by $\frac{s_p(1-\tau^c)-z}{\tau^c z}$. where K_g is the capital stock owned by the government, K_p is that of the private sector. For $z<z^*$, $k = f'^{-1}\left(\frac{n}{s_p(1-\tau^c)}\right)$. For a fixed τ^c, changes in z have no effect on the wages received by workers. The payments from the government (per worker) are $(1-z)r(k-(1-\tau^c)k_p)$. We already noted that at the limiting case where $z=0$, workers are worse off than they would be without taxation.
73. That is, the equilibrium is described by the solution to the pair of equations (in the natural notation):
    (i) $(1-\tau^c)s_p f_{k_p}=n$
    (ii) $\frac{\tau^c f_{k_p} k_p}{k_g} + f_{k_g} = n$
74. An earlier version of the ideas in this section were delivered as a keynote address at the National Tax Association annual meetings, Santa Fe, November, 2014.
75. See Stiglitz (2015).
76. This is particularly relevant given the literature which has suggested that the *pure* returns to capital should be taxed at a zero rate, based on a misinterpretation of the Atkinson-Stiglitz (1976) result. See also Stiglitz (2015).
77. We cited evidence that that was the case earlier.
78. This particular formulation has the characteristic of a jump in the level of savings. A formulation with similar consequences is $s(W)=s_0$ for $W \le W_1$; $s(W)W = s_0W_1+s_1(W-W_1)$ for $W_1 \le W \le W_2$; and $s(W)W=s_0W_1+s_1(W_2-W_1)+s_2(W-W_2)$ for $W \ge W_2$, with $s_1 >> s_0$ and $s_1 >> s_2$.
79. In Stiglitz (2015b), we also suggest that that model also could not adequately explain the growth of wealth inequality that has been observed.
80. We also noted in Part I that there has been an increase in other forms of rents, and when capitalized, these too give rise to an increase in wealth.
81. This results should be contrasted with that of Part III of this paper. The difference arises from the difference in the determinants of savings. We believe that the assumptions made here provide a better description of today's economy.
82. Arnott and Stiglitz (1978) and Stiglitz (2015c) have precisely calculated urban land values for cities of different geometries, relating it to aggregate transport costs and spending on local public goods.
83. The value of land is $(1-t^L)\frac{f_{T^\wedge}}{f_k}$. The reduction in f_k will normally partially offset the tax, so that the value of land will not go down commensurately with the reduction in $1-t^L$.
84. For a more complete analysis of this model, see Stiglitz (2010b). Similar results hold with money, rather than land, as we show in the Part IV of this paper.
85. The other interesting case is that where *land as an unproductive store of value*.

    If $n=0$ and $s_p=1$, then in steady state, the interest rate will be zero, and the price of land will be constant. (4.6) takes on the form

$$k_{t+1}+q_{t+1}T=k_t+q_tT. \tag{4.9}$$

It should be clear that k^* in combination with any value of q is an equilibrium: as before, the value of land is indeterminate.

On the other hand, if $s_p<1$, the analysis of the steady state presents some problems. Assume that there were a steady state. r^* will be positive, and that means that the price of land has to be ever increasing – but that in turn would imply that wealth is increasing and capital is an increasingly diminishing fraction of wealth. And who would hold this ever increasing wealth?

The only value of q_0 consistent with the equilibrium conditions is $q=0$. If q were ever to be positive, *for the capital arbitrage equation to be satisfied,* an increasing fraction of savings has to be devoted to holding land, and a diminishing amount goes into capital accumulation. The rate of interest would, accordingly, rise. But as that happens, capital gains increase even more, diverting even more savings into land. In short, as before, the equilibrium (with $q=0$) is not stable.

86. If R is the rent from the land, and r is the real interest rate, then the value of land $V_T=R/r$, so that there is an equiproportionate increase in the value of land from a decrease in the real interest rate.
87. This analysis applies to a comparison across steady states with different K.
88. $p=\boldsymbol{M}(W,p,u)$, and, assuming that expectations about capital gains are fixed, $\frac{\partial \mathrm{W}}{\partial \tau^L}=\frac{\partial \mathrm{p}}{\partial \tau^L}=\frac{pM_u}{1-M_w-M_p}\cdot$ A natural stability condition ensures that the denominator is positive. Since $M_u<0$, the tax reduces the price of land.
89. But there are important effects going the other way, and which almost surely predominate – for instance, the increased insecurity that the non-rich face, not adequately reflected in income statistics.
90. The increase in the price of land is only partially explained by the discussion of this section. Section 4 argues that the expansion of the credit supply provides an important part of the explanation.
91. The recurrence of bubbles has been noted by Kindleberger (1978).
92. Hahn (1966), Shell and Stiglitz (1967).
93. For simplicity, we assume that $\mathrm{F_K}$ approaches infinity as K approaches zero, and that the marginal product of capital falls to zero only as K approaches infinity.
94. In the Kaldor model, $r=n/s$ where here, s is the savings rate of capitalists; in the Solow model, where everyone has the same rate, $r<f/k=n/s=0$. Similar results obtain in the two-class model of Part III of this paper.
95. See, e.g. Stiglitz (2014) and the references cited there.
96. Similar results hold for a model with money, such as that formulated in section 4.
97. Similar results can be obtained if we assume savings are a fixed fraction of overall income (including capital gains).
98. If $\mathrm{s_p}>0$, there is a unique solution to (7.4a) and (7.8).
99. The dynamics are oscillatory if $(s_K(K^*,p^*,0)-p^*F_{KK}(K^*))^2<4(-s_p(K^*,p^*,0)p^*F_{KK}(K^*))$.
100. Note that $\left.\frac{dp}{dK}\right|_{\frac{dK}{dt}=0}=-\frac{s_K-pF_{KK}(1-s_3)}{s_p+(F_K-\mu)s_3}\cdot$ If as K gets small, s_p remains greater than $(1-s_3)(F_K-\mu)$, then the $\left.\frac{dp}{dK}\right|_{\frac{dK}{dt}=0}$ locus will hit the vertical axis. (s_3 is the (marginal) savings out of capital gains. It is natural to assume that $0<s_3<1$.

Along any trajectory, $\frac{dp}{dK}=\frac{dp/dt}{dK/dt}=\frac{p(F_K-\mu)}{s-p(F_K-\mu)}$ which goes to zero as p goes to zero.

101. The steady state can also be described by the intersection of (7.13) and the locus $\frac{sf(k^*) - nk^*}{(1-s)n} = q^*$.
which gives the values of k and q such that $dk/dt=0$ *when* $\frac{d}{dt}(\log p) = n$.
102. A sufficient condition for this is that land and capital are complements.
103. If there is more than one solution, k^{**} is defined as the smallest.
104. When the price is too low, eventually, the price may shrink to zero. For the rest of the analysis, we ignore this case.
105. The sign depends on whether for the $dk/dt=0$ locus, $\frac{\partial \mathrm{k}}{\partial(1-t^{cg})}$, conditional on fixed q, is greater than for the $dq/dt=0$ locus, i.e. whether at q^*, $(1-s)(f_{T\wedge k}-qf_{kk})$ is greater or less than $sf'(k^*)-n$. Either seems possible.
106. The model is obviously stylized, but there are good reasons why land should serve better as collateral than capital goods – capital goods tend to be constructed for specific purposes, and are less malleable, less alterable to other uses, with often large asymmetries of information concerning the prospects of returns not only in the intended use, but also in alternative uses. There are other reasons that the provision of credit typically gets reflected in land bubbles (or bubbles in other fixed assets): when the price of capital goods exceeds the production costs, the supply will increase, and this limits the extent to which the price can rise or the duration of any bubble associated with a produced good. (Nonetheless, bubbles of produced goods do occur – the tech bubble in the nineties and the tulip bubble in the seventeenth century being the most famous instances.)

    The model can easily be generalized. We have assumed, in particular, that capitalists-entrepreneurs are the only ones who do *real* savings, while landowners/rentiers simply buy land, and that credit is only provided to the latter rather than the former. In the final subsection, we allow credit against capital goods as collateral.
107. For simplicity, here we assume that s_p is the gross savings rate, which is assumed to be fixed and based on gross income, where r is now the gross return to capital. We could rewrite all of these equations based on net savings and net income, without changing any of the results.
108. Because we do not want to address issues involving the banking system and the wealth of its owners, we will simplify the analysis and assume that it is government owned. As formulated, the banking system makes neither profits nor losses.
109. In Part I of this paper, we noted that this characterized several countries.
110. This analysis, however, does not explain why workers' compensation should have decreased even as average productivity has increased. Of course, average productivity could have increased even if the ratio of capital per effective labor unit decreased, simply because of technological change.
111. In the analysis below, we assume that the rate charged is zero. This is a simplifying assumption. All that is required is that the rate charged by less than F_K.
112. This assumes, of course, that the change in policy was not anticipated.
113. This can be seen most transparently in a situation where the economy is initially at full employment. Assume that savings (consumption) is interest insensitive. If financial regulations were eased, so that banks could lend more, given their deposits and net worth (reserve and capital adequacy requirements were loosened), it would appear that banks could lend more, and if banking is profitable at the margin, each bank would believe such a policy would be desirable. But if they all

started to lend more, there would be excess demand, and the Fed would have to raise interest rates, to tighten credit in a fully offsetting way.

114. We again assume a constant labor supply and normalize the labor supply at unity.
115. With all of profits going into (gross) investment, aggregate consumption must equal wages. Second period consumption is just $B+r_gB$, that is, $C_1+C_2=(1-s)(w-\tau)+\tau+B=w-s(w-\tau)+B=w$, from which the result follows immediately.
116. Recall that capitalists' savings behavior determines r: $s_p r=n$. In the remainder of this section, we assume $s_p=1$.
117. $\tau=r^*B=sr_gw/(1+sr_g)$. $\dfrac{dlog\tau}{dlogr_g}=1-\dfrac{sr_g}{1+sr_g}=\dfrac{1}{1+sr_g}>0.$
118. $C_2=(1+r_g)B=(1+r_g)\dfrac{sw}{1+sr_g}$. $\dfrac{dlogc_2}{dlogr_g}=\dfrac{r_g}{1+r_g}-\dfrac{sr_g}{1+sr_g}=\dfrac{r_g(1+sr_g-s-sr_g)}{(1+r_g)(1+sr_g)}=\dfrac{r_g(1-s)}{(1+r_g)(1+sr_g)}>0.$
119. Steady-state utility of workers is maximized at $U(C_1,w-C_1)$, i.e. where $U_1=U_2$. Individuals will choose this allocation if $r=0$. One could conduct a full dynamic analysis, rather than focusing on steady states, with much the same results. Focusing on steady states greatly simplifies the calculations.
120. If we had expanded the model to include land (as in earlier sections), there will also be an increase in its value.
121. In our model, the rate of growth of the labor force is zero, and the rate of labor augmenting technical progress is zero. Thus, the long run rate of growth of the economy is zero. The critical condition involves the relationship between the rate of interest and the rate of growth.

    Standard focuses on the zero lower bound constraint. This is a lower bound on the nominal interest rate. In the United States, in the aftermath of the crisis, the real interest rate has been negative.
122. $\dfrac{dlog(sw/K)}{dlog(r_g)}=\left(\dfrac{-K^2F_{KK}}{f-KF_K}-1\right)\dfrac{dlog(K)}{dlog(r_g)}=\left(\dfrac{S_K}{\varepsilon}-1\right)\dfrac{dlog(K)}{dlog(r_g)}$ where ε is the elasticity of substitution and S_K is the share of capital. We note that because we have normalized labor supply at unity, which is fixed, the capital–labor ratio, usually denoted by k, is the same as the level of capital stock, K (The elasticity of substitution is equal to $F_K(F-KF_K)/KF_{KK}F$).
123. For a discussion of the merits of income contingent loans, see Chapman *et al* 2014.
124. We note that we are able to derive a simple formula describing tail inequality. In the case of a Solow model (all save the same fraction s of their income) with all receiving the same wage but stochastic returns to capital, the Pareto coefficient is given by $2s^2\dfrac{1-S_k}{S_k^2n\bar{\sigma}^2}$, where $\bar{\sigma}^2$ is the variance of returns. In the limiting case where variance is zero, we obtain the earlier result of Stiglitz (1969) that there is no inequality. Note again that the difference between r and growth plays no role, but the share of capital does.
125. We show that increases in credit available (decreases in collateral requirements) can give rise to increases in land values, but we have also shown that there can be land bubbles even in the absence of credit expansion (though recent bubbles have clearly been supported by such credit expansion.)
126. See, for example, Piketty (2014) and Stiglitz (2012b). Such changes affect both the distribution of income and wealth at any moment of time as well as the dynamics that describe the evolution of those variables. This paper has taken technology as exogenous, but as Braverman and Stiglitz (1989) point out, technology and technological change itself is affected by societal inequalities. Sharecropping is a prevalent

tenancy arrangements in economies with large disparities in land ownership, but not otherwise. But the choice of technology at one moment affects the distribution of income and wealth and wealth dynamics, and even the nature of technological change (Greenwald and Stiglitz, 2014).

127. The points raised here (and similar points made elsewhere in this paper) are echoed in Suresh Naidu's excellent review of Piketty (2014).
128. The result follows immediately upon observing that we can write $\bar{w}(K)=(F_{L_1}\omega+F_{L_2}(1-\omega))$, and treating K and ω as functions of time.
129. Now $\frac{k^{w*}}{k^*}=\frac{ns\left(\frac{1-S_k}{S_k}+\tau^c\right)}{s_p(1-\tau^c)+s(k^*)\tau^c\hat{n}}$. So long as $s_p>ns$, the direct effect of an increase in taxes is to increase the importance of life cycle savings. If the elasticity of substitution is greater than one, the indirect effect is also positive, so long as s'≤0. (Now the workers' savings rate plausibly depends on k, since there is no taxation on the return to life cycle savings, and the before tax return increases.)
130. We can in principle derive the savings functions from V.
131. We have made use of the fact that for an indirect utility function, $\frac{\partial V}{\partial r}=s(k^*)w(k^*)\frac{\partial V}{\partial Y^W}$.
132. This analysis assumes that social welfare is only assessed from the perspective of workers (who receive no inheritances.) It ignores the welfare of the capitalists. If their wellbeing were also included within the social welfare function, the optimal tax would obviously be different. Note the steady state income of the capitalists always decreases with taxation, that is, $\frac{d}{d\tau^c}((1-\tau^c)rk^*)=(1-\tau^c)(f'(k^*)+f''(k^*)k^*)\frac{dk^*}{d\tau^c}-r(k^*)k^*=\frac{f'(k)f'(k)}{f''(k)}<0$, but so does income per capita.

References

Acemoglu, Daron, D. Autor, D. Dorn, G. Hanson, and B. Price (2014) "Import Competition and the Great U.S. Employment Saga of the 2000s," NBER Working Paper.

Aitchison, John and J.A.C Brown (1957) "The Lognormal Distribution," *The Economic Journal*, vol. 67, no. 268, pp. 713–15.

Arnott, R. and J.E. Stiglitz (1979) "Aggregate Land Rents, Expenditure on Public Goods and Optimal City Size," *Quarterly Journal of Economics*, vol. 93, no. 4, pp. 471–500.

Arrow, Kenneth J., Hollis B. Chenery, Bagicha S. Minhas, and Robert M. Solow (1961) "Capital–Labor Substitution and Economic Efficiency," *The Review of Economics and Statistics*, vol. 43, no. 3, pp. 225–50.

Atkinson, A.B. and J.E. Stiglitz (1976) "The Design of Tax Structure: Direct Versus Indirect Taxation," *Journal of Public Economics*, vol. 6 (July–August), pp. 55–75.

Autor, David (2002) "Skill Biased Technical Change and Rising Inequality: What is the Evidence? What are the Alternatives?" Working paper, MIT and NBER.

Autor, D.H., L.F. Katz, and M.S. Kearney (2008) "Trends in US Wage Inequality: Revising the Revisionists," *Review of Economics and Statistics,* vol. XC, no. 2, pp. 300–23.

Banerjee, Abhijit and Sendhil Mullainathan (2010) "The Shape of Temptation: Implications for the Economic Lives of the Poor," NBER Working Paper 15973.

Becker, Gary S. and Nigel Tomes (1979) "An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility," *Journal of Political Economy*, vol. 87, no. 6, pp. 1153–89.

Becker, Gary S. and Nigel Tomes (1986) "Human Capital and the Rise and Fall of Families," *Journal of Labor Economics,* vol. 4, no. 3, pp. S1–S39.

Becker, Gary S. and Nigel Tomes (1994) "Human Capital and the Rise and Fall of Families," in Gary S. Becker (ed.), *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education*, 3rd edition. Chicago: The University of Chicago Press.

Bernanke, Ben S. (2005) "The Global Saving Glut and the US Current Account Deficit," *The Sandridge Lecture*, Virginia Association of Economists, Richmond, VA.

Bevan, D. and J.E. Stiglitz (1979) "Intergenerational Transfers and Inequality," *The Greek Economic Review*, no. 1, pp. 8–26.

Bowles, Samuel and Herbert Gintis (2002) "The Inheritance of Inequality,." *Journal of Economic Perspectives*, vol. 16, no. 3, pp. 3–30.

Braverman, A. and J.E. Stiglitz (1989) "Credit Rationing, Tenancy, Productivity and the Dynamics of Inequality," in P. Bardhan (ed.), *The Economic Theory of Agrarian Institutions*. Oxford: Clarendon Press, pp. 185–201.

Card, David and John E. DiNardo (2002) "Skill-Biased Technological Change and Rising Wage Inequality: Some Problems and Puzzles," *Journal of Labor Economics*, vol. 20, no. 4, pp. 733–83.

Carroll, Christopher D. (1998) "Why Do the Rich Save So Much?," NBER Working Papers 6549.

Champernowne, D.G. (1953) "A Model of Income Distribution," *Economic Journal,* vol. LXIIII, pp. 318–51.

Chapman, B., T. Higgins, and J. E. Stiglitz (eds) (2014). *Income Contingent Loans: Theory, Practice and Prospects*. Basingstoke, UK and New York: Palgrave Macmillan.

Dasgupta, Partha and Joseph Stiglitz (1980) "Industrial Structure and the Nature of Innovative Activity," The Economic Journal, vol. 90, no. 358, pp. 266–93.

Domar, Evsey D., and Richard A. Musgrave (1944) "Proportional Income Taxation and Risk-taking," *Quarterly Journal of Economics*, vol. 58, no. 3, pp. 388–422.

Dynan, Karen, Jonathan Skinner, and Stephen P. Zeldes (2004) "Do the Rich Save More?," *Journal of Political Economy*, vol. 112, no. 2, pp. 397–444.

Federal Reserve Board (2015) *The State of Lending in America & Its Impact on U.S. Households: The Cumulative Costs of Predatory Practices*.

Feldstein, Martin (2014) "Piketty's Numbers Don't Add Up," *Wall Street Journal*, May 14. Available at http://online.wsj.com/articles/SB10001424052702304018045795576641 76917086.

Galbraith, James Kenneth (2012) *Inequality and Instability: A Study of the World Economy Just Before the Great Crisis*. Oxfotrd: Oxford University Press.

George, Henry (1879) *Progress and Poverty: An Inquiry into the Cause of Industrial Depressions and of Increase of Want with Increase of Wealth* VI. New York: Robert Schalkenbach Foundation.

Giovannoni, Olivier G. (2014) "What Do We Know About the Labor Share and the Profit Share? – Part III: Measures and Structural Factors," Bard College, March 3.

Giovannoni, Olivier G. (2015) "Inequality: Challenge of the Century?," Presentation to the ASSA meetings, Boston, January 3.

Griliches, Zvi (1969) "Capital-Skill Complementarity," *The Review of Economics and Statistics*, vol. 51, no. 4, pp. 465–8.

Greenwald, Bruce and Judd Kahn (2009) *Globalization: n. the Irrational Fear That Someone in China Will Take Your Job*. Hoboken, NJ: John Wiley & Sons.

Guzman, M. and J.E. Stiglitz (2014) "Pseudo-wealth and Consumption Fluctuations," Columbia University.
Hahn F. (1966) "Equilibrium Dynamics with Heterogeneous Capital Goods," *Quarterly Journal of Economics*, vol. 80, pp. 633–46.
Henry, C. and J.E. Stiglitz (2010) "Intellecutal Property, Dissemination of Innovation, and Sustainable Development," *Global Policy*, vol. 1, no. 1, October, pp. 237–51.
Hotelling, Harold (1931) "The Economics of Exhaustible Resources," *Journal of Political Economy,* April.
Jones, C., and Paul Romer (2009) "The New Kaldor Facts: Ideas, Institutions, Population, and Human Capital," NBER Working Paper 15094, June.
Kaldor, Nicholas (1957) "A Model of Economic Growth," *The Economic Journal*, vol. 67, no. 268, pp. 591–624.
Kaldor, Nicholas (1961) "Capital Accumulation and Economic Growth," in F.A. Lutz and D.C. Hague (eds), *The Theory of Capital*. St. Martins Press, pp. 177–222.
Katz, Michael L. and Carl Shapiro (1994) "Systems Competition and Network Effects," *Journal of Economic Perspectives*, vol. 8, no. 2, pp. 93–115.
Kindleberger, Charles P. and Robert Aliber (1978) *Manias, Panics, and Crashes: A History of Financial Crisis*. New York: John Wiley & Sons.
Knolly, Katharina, Moritz Schularickz, and Thomas Steger (2014) "Home Prices Since 1870: No Price Like Home," *Vox*, November 1.
Knolly, Katharina, Moritz Schularickz, and Thomas Steger (2015) "No Price Like Home: Global House Prices, 1870–2012," working paper.
Krusell, Per, Lee E. Ohanian, Jose-Victor Rios-Rull, and Giovanni L. Violante (2000) "Capital-Skill Complementarity and Inequality: A Macroeconomic Analysis," *Econometrica*, vol. 68, no. 5, pp. 1029–54.
Kuznets, Simon (1955) "Economic Growth and Income Inequality," *The American Economic Review*, vol. 45, no. 1, pp. 1–28.
Lebergott, Stanley (1959) "The Shape of the Income Distribution," *The American Economic Review*, vol. 49, no. 3, pp. 328–47.
Mallick, Debdulal (2007) "The Role of the Elasticity of Substitution in Economic Growth: A Cross-Country Test of the de La Grandville Hypothesis," *Economic series*, Deakin University, Faculty of Business and Law, School of Accounting, Economics and Finance.
Mason, Joshua W. (2014) "Three Essays in Macroeconomic History", University of Massachusetts Amherst.
Mason, Joshua W. (2015) "Disgorge the Cash," Working Paper, Roosevelt Institute.
Milevsky, Moshe A, and Huaxiong Huang (2011) "Spending Retirement on Planet Vulcan: The Impact of Longevity Risk Aversion on Optimal Withdrawal Rates," *Financial Analysts Journal*, vol. 67, no. 2, pp. 45–58.
Naidu, Suresh (2014). "Capital Eats the World." https://www.jacobinmag.com/2014/05/capital-eats-the-world/.
OECD (2011) *Divided We Stand: Why Inequality Keeps Rising*. Paris: OECD.
OECD (2013) *Pension Markets in Focus*. Paris: OECD.
Pasinetti, L. (1962) "The Rate of Profit and Income Distribution in Relation to the Rate of Economic Growth," *Review of Economic Studies*, vol. 29, no. 4, pp. 267–79.
Piketty, Thomas (2014) *Capital in the Twenty-First Century*. Cambridge, MA: The Belknap Press of Harvard University Press.
Piketty, Thomas and Gabriel Zucman (2014) "Capital is Back: Wealth–Income Ratios in Rich Countries 1700–2010," *Quarterly Journal of Economics*, vol. 129, no. 3, pp. 1155–210.
Piketty, Thomas and Emmanuel Saez (2003) "Income Inequality in the United States, 1913–1998," *Quarterly Journal of Economics*, vol. 118, no. 1, pp. 1–39. (Longer updated

version published in A.B. Atkinson and T. Piketty eds, Oxford University Press, 2007.) (Tables and figures updated to 2012 in Excel format, September 2013.)
Piketty, Thomas, Emmanuel Saez, and Stefanie Stantcheva (2014) "Optimal Taxation of Top Labor Incomes: A Tale of Three Elasticities," *American Economic Journal: Economic Policy*, vol. 6, no. 1, pp. 230–71.
Pogge, Thomas and Sanjay Reddy (2010) "How Not to Count the Poor," in Joseph Stiglitz, Sudhir Anand and Paul Segal (eds), *Debates in the Measurement of Poverty*. Oxford: Oxford University Press.
Roberts, Sam (2014) "Why are Rockefellers Moving from 30 Rock? 'We Got a Great Deal'," *New York Times*, November 24, p. A1.
Saez, Emmanuel and Gabriel Zucman (2014) "Wealth Inequality in the Unites States Since 1913: Evidence from Capitalized Income Tax Data," NBER Working Paper 20625.
Scheinkman, J.A., and W. Xiong (2003) "Overconfidence and Speculative Bubbles," Journal of Political Economy, vol. 111, no. 6, pp. 1183–219.
Shell, Karl, and Joseph E. Stiglitz (1967) "Allocation of Investment in a Dynamic Economy," *Quarterly Journal of Economics*, vol. 81, pp. 592–609.
Shell, K., M. Sidrauski, and J.E. Stiglitz (1969) "Capital Gains, Income and Savings," *Review of Economic Studies*, vol. 36, no. 1, pp. 15–26.
Shierholz, Heidi and Lawrence Mishel (2013) "A Decade of Flat Wages," Economic Policy Insitute Report, available at http://www.epi.org/publication/a-decade-of-flat-wages-the-key-barrier-to-shared-prosperity-and-a-rising-middle-class/.
Shierholz, Heidi, Lawrence Mishel, and John Schmitt (2013) "Don't Blame the Robots: Assessing the Job Polarization Explanation of Growing Wage Inequality," Economic Policy Institute, November 19.
Sraffa, Piero (1960) *Production of Commodities by Means of Commodities: Prelude to a Critique of Economic Theory*. London: Cambridge University Press.
Stiglitz, Joseph E. (1966) "The Distribution of Income and Wealth Among Individuals," Presented at the December 1966 meetings of the Econometric Society, San Francisco, December 29 (MIT, mimeo).
Stiglitz, Joseph E. (1969a) "Distribution of Income and Wealth Among Individuals," *Econometrica*, vol. 37, no. 3, pp. 382–97.
Stiglitz, Joseph E. (1969b) "The Effects of Income, Wealth and Capital Gains Taxation on Risk-Taking," *Quarterly Journal of Economics*, vol. 83, no. 2, pp. 263–83.
Stiglitz, Joseph E. (1974) "The Cambridge–Cambridge Controversy in the Theory of Capital: a view from New Haven: a Review Article," *Journal of Political Economy*, vol. 82, no. 4, pp. 893–903.
Stiglitz, Joseph E. (1975) "The Theory of Screening, Education and the Distribution of Income," *American Economic Review*, vol. 65, no. 3, June, pp. 283–300. Reprinted in *Selected Works of Joseph E. Stiglitz, Volume I: Information and Economic Analysis*. Oxford: Oxford University Press, 2009, pp. 99–121.
Stiglitz, Joseph E. (2003) *The Roaring Nineties: A New History of the World's Most Prosperous Decade*. New York: W.W. Norton & Company.
Stiglitz, Joseph E. (2010a) *Freefall: America, Free Markets, and the Sinking of the World Economy*. New York: WW Norton.
Stiglitz, Joseph E. (2010b) "The Wobbly Economy," mimeo, Columbia University.
Stiglitz, Joseph E. (2012a) *The Price of Inequality: How Today's Divided Society Endangers Our Future*. New York: W.W. Norton. Also published in the UK by Penguin/Allen Lane. Translated in simplified Chinese by China.
Stiglitz, Joseph E. (2012b) "Macroeconomic Fluctuations, Inequality, and Human Development," *Journal of Human Development and Capabilities*, vol. 13, no. 1, pp. 31–58.

Reprinted in Deepak Nayyar (ed.), *Macroeconomics and Human Development*. London: Taylor and Francis.
Stiglitz, Joseph E. (2013) "Stable Growth in an Era of Crises: Learning from Economic Theory and History," *Ekonomi-tek,* vol. 2, no. 1, pp. 1–38. (Originally delivered as keynote lecture to the Turkish Economic Association, Izmir, November, 2012.)
Stiglitz, Joseph E. (2014a) "Unemployment and Innovation," NBER Working Paper 20670.
Stiglitz, Joseph E. (2014b) "Inequality in America: A Policy Agenda for a Stronger Future: 2014, Daniel Patrick Moynihan Lecture on Social Science and Public Policy," *Annals of the American Academy of Political and Social Scieinces*.
Stiglitz, Joseph E. (2014c) "Tapping the Brakes: Are Less Active Markets Safer and Better for the Economy?" Presented at the Federal Reserve Bank of Atlanta 2014 Financial Markets Conference Tuning Financial Regulation for Stability and Efficiency, April 15. Available at http://www.frbatlanta.org/documents/news/conferences/14fmc/Stiglitz.pdf.
Stiglitz, Joseph E. (2015a) "Fed Policy, Inequality, and Equality of Opportunity," keynote address to The Ninth Biennial Federal Reserve System Community Development Research Conference, April 3.
Stiglitz, Joseph E. (2015b) "New Theoretical Perspectives on the Distribution of Income and Wealth among Individuals," Roosevelt Institute Working Paper.
Stiglitz, Joseph E. (2015c) "Devolution, Independence, and the Optimal Provision of Public Goods," *Economics of* Transportation, vol. 4, no. 1, pp. 82–94.
Stiglitz, Joseph E. (2015d) "In Praise of Frank Ramsey's Contribution to the Theory of Taxation," *Economic Journal*.
Stiglitz, Joseph E. and Bruce. C. Greenwald (2014) *Creating a Learning Society: A New Approach to Growth, Development, and Social Progress*. New York: Columbia University Press.
Stiglitz, Joseph E. and Andrew Weiss (1981) "Credit Rationing in Markets with Imperfect Information," *American Economic Review*, vol. 71, no. 3, pp. 393–410.
Stiglitz, Joseph E., A. Sen, and J-P Fitoussi (2010) *Mismeasuring Our Lives: Why GDP Doesn't Add Up*. New York: The New Press.
United Nations (2012) *World Population Prospects: The 2012 Revision*.
Wolff, Edward N. and Maury Gittleman (2011) "Inheritances and the Distribution of Wealth Or Whatever Happened to the Great Inheritance Boom?," BLS Working Paper 445, January.
Young, Andrew T. (2013) "US Elasticity of Substitution and Factor-Augmentation at the Industry Level," *Macroeconomic Dynamics*, vol. 17, no. 4, pp. 861–97.

2

Reflections on the "Equity and Development" World Development Report Ten Years Later*

François Bourguignon
Paris School of Economics

It is fair to say that in the mind of its instigators 'inequality' rather than 'equity' was to be the dominant theme of the 2006 World Development Report. Its goal was to challenge the dominant view, in the World Bank and elsewhere in the development community, that, in their aim to reduce and possibly eradicate poverty, development strategies should focus mostly on aggregate growth. The main idea to be developed in the report was that the whole distribution of income within the population, rather than only its mean, mattered and should be a major concern for policy makers. Namely, the degree of inequality of the income distribution affected poverty reduction in two ways. Not only was it reducing the share of the gain from growth actually accruing to the poorest, but it was also slowing down growth itself.

Inequality was then the focus of a rather hot policy debate. On one side were those who thought it should be ignored and poverty reduction should indeed rely exclusively on the pursuit of aggregate growth. Quite influential in this respect had been the "Growth is good for the poor" paper by Dollar and Kraay (2001), which suggested that, on average across country experiences, growth benefits in the same proportion to the rich, the middle class and the poor. On the other side of the debate were those who were stressing that the impact of growth on poverty depended on both the growth of total income and its distribution, so that growth was not necessarily 'pro-poor'.[1]

There was also the theoretical literature of the 1990s, summarized in Aghion et al. (1999), showing that the issue was not so much whether inequality was increasing or decreasing with growth but whether an excessive level of

*The main ideas in this paper were first presented at the workshop on "Inequality and Development" jointly organized by the International Economic Association and the World Bank in June 2014 at the Dead Sea in Jordan. I thank the participants to the workshop for helpful comments with special thanks to Francisco Ferreira, Martin Ravallion and Michael Walton who commented on a previous draft of this paper.

inequality could reduce economic efficiency and slow down growth. Early evidence based on cross-country comparisons seemed to confirm such a negative effect of inequality on growth, even though it lacked robustness and was leaving considerable room for difference across country experiences.

Within such a context, the WDR 2006 on 'equity and development' tried to set the issue of inequality at the centre of the development debate. However, in view of today's new emphasis on inequality issues and the success of the recent books by Stiglitz (2012) and Piketty (2014) the WDR 2006 may appear a bit shy. Why did it refer to 'equity', apparently a less politically loaded word than 'inequality'? Why was only limited space devoted to income redistribution and the taxation of top incomes? Why did the report end on such soft recommendations as 'leveling the playing field' and 'improving governance' rather than more egalitarian messages?

There are two answers to the preceding questions. The first one has to do with diplomacy and the second one is analytical.

Diplomacy was required in an institution where the word inequality was still seldom used in official communication and the reduction of poverty was seen as essentially requiring faster growth without any attention being paid to the distribution of its product. It thus seemed that dealing with the issue of inequality and development under the heading of 'equity' would avoid harsh reactions from those who considered in those days that talking about inequality necessarily obeyed a strongly egalitarian view of the world.

But there was more than diplomacy. While planning the work for this report, it became clear in the mind of the authors that the theoretical economic literature on inequality and growth mentioned above had most often been misread. Its implication was not that reducing the inequality of *incomes*, for instance through straight income redistribution, would accelerate growth, as several leading papers in that literature had been interpreted. It was that market failures responsible for an unequal access of people to income-generating facilities like credit, education, insurance or justice, as well as the confiscation of political power by a predatory elite, could cause simultaneously an unequal distribution of income, a high level of poverty and slower growth. The WDR 2006 was thus to put the emphasis on the distribution of *opportunities* and on the policies to make it more equal. Referring to 'equity' rather than 'inequality' was meant to signal this shift of emphasis from the distribution of income, or outcomes in general, to opportunities. It was also meant to stress that the relationship between inequality and development goes very much beyond that between the Gini coefficient of income inequality and the rate of growth of GDP, as summarily assumed in much of the empirical literature of those days. In the perspective of the then ongoing debate on poverty, inequality and economic development, this plea for returning to a truly structural approach to these issues was indeed essential and did much to get some consensus around

the main messages of the report. At the same time, it is possible that, by pushing the analysis in various alternative directions, all the implications of this approach in terms of 'inequality' of incomes were not fully analyzed or properly emphasized, as I will discuss in this chapter.

The WDR 2006 was well received and its contribution to changing the mindset in part of the development community duly recognized. Inequality now is fully part of the debate on development and development policies. The Gini coefficient is increasingly among the indicators by which development performances are being judged, along with GDP growth or the $1.25 a day poverty headcount. It has been proposed that inequality should be part of the Sustainable Development Goals which are to replace the Millennium Development Goals beyond 2015. The 'shared prosperity' goal recently introduced by the World Bank – that is, the growth of the mean income of the 40% poorest people – may itself be considered as involving an inequality criterion when explicitly compared to progress in GDP per capita. Of course, all this is not due to the WDR 2006, but it has most probably been somewhat instrumental in that evolution, even though this extension of standard development criteria refers to the inequality of outcomes rather than to 'equity' and the distribution of opportunities, as advocated in the report.

As the chief economist of the World Bank in those days, I have been the initiator and then the overall supervisor of the report. Rereading it, I still find today it is a thoughtful and helpful contribution to development thinking and policy, and I am grateful to the authors for their excellent and important work.[2] I must say I even found it more complete than what I remembered. Yet it is true that some issues may not have been given the emphasis they deserved and may have made the message of the report weaker than it should have been. This is the case, in particular, with the issue of income inequality and top incomes or with taxation and market regulation policies that would redistribute income or wealth from the top to the middle and the bottom of the income ladder.

This chapter revisits the WDR 2006 and comes back to this issue of the potential role of income redistribution within a development policy framework that focuses on both equity and efficiency. It tries to bring to the forefront several points which may not have been given enough prominence in the WDR 2006 and make clearer that part of its messages that refer to income inequality rather than the inequality of opportunities.

The chapter is divided into five parts. The first one is devoted to a very brief summary of the main messages of the WDR 2006, essentially the complementarity between equity and efficiency, and the arguments behind them. Then four key issues are discussed. The definition of the concepts of equity and efficiency used in the report and their possible limitations, noted by some reviewers, are discussed in section 2.2. The mostly microeconomic evidence about the relationship between efficiency and equity and the difficulty of aggregating it

up at the macro level are analyzed in section 2.3. The issue of income redistribution and its implicit but central role in the main policy implications of the report is handled in section 2.4. Finally, the last section is devoted to the issue of the 'top incomes' and the importance to be given to income inequality per se in the development policy debate. The main points in the chapter are then summarized in a brief conclusion.

2.1 The complementarity between equity and efficiency and the main messages of the WDR 2006

The main message of the WDR 2006 can be summarized as follows: 'Equity matters both intrinsically and instrumentally; in particular, it is the inequality of opportunity, and not necessarily the inequality of outcomes (e.g. incomes) that hinders growth and poverty reduction.' The following quotes from the Overview of the report illustrate this message:

i. *Public action should focus on the distributions of assets, economic opportunities, and political voice, rather than directly on inequality in incomes (p. 3)*
ii. (The report) "*presents evidence that the inequality of opportunity is inimical to sustainable development and poverty reduction.*" *(p. 3)*
iii. (It derives) "*policy implications that center on the broad concept of leveling the playing field -both politically and economically."(p. 3)*
iv. "*It makes the case for investing in people, expanding access to justice, land, and infrastructure, and promoting fairness in markets.*" *(p. 4)*

The first two quotes convey the idea that it is through the inequality of opportunity that the inequality of income can be corrected and that it is mostly through opportunities that inequality has a negative impact on economic growth and poverty reduction. Hence the recommendation follows to level the playing field and to guarantee equal access to education of equal quality, credit or infrastructure to all citizens. This equalizing of opportunities also had to address the decision making institutions and the danger they would be captured by a predatory elite.

The argument in support of these messages and recommendations was directly inspired from the theoretical literature on inequality and growth, and in particular the contributions by Galor and Zeira (1993) and Aghion and Bolton (1996) which investigated the implications of the inequality of wealth and credit market imperfections on economic efficiency and growth. The argument may be summarized as follows.

Assume that, due to information asymmetry on the probability of success of the projects to be financed thanks to their loans, banks require from their borrowers a collateral amounting to X. It follows that people whose wealth is

below X cannot undertake projects with a fixed cost above X, even though their rate of return may be higher than the rate of interest in the economy. Under these conditions, transferring wealth from people with wealth much above *X* to people just below *X* would allow a number of the latter to access credit and to invest in their high-return projects. As the return on these projects is higher than the rate of interest, total income would increase.

This argument is illustrated in Figure 2.1 where a wealth transfer, *x*, is made between individual *R* and individual P whose wealth is just below the borrowing threshold, X. The size of the transfer is such that it allows P's wealth to jump above X, so that P can undertake a project with a net annualized return, *b*. On the side of *R*, transferring *x* is equivalent to an annualized loss of income equal to $a=r\,x$, where *r* is the rate of interest. As the project undertaken by P has a rate of return above *r,* it follows that $b>a$ and the aggregate income in the economy rises by $b-a\ (>0)$. Equalizing the wealth distribution through a transfer from a 'rich' person R to a 'poor' person P, just below the borrowing threshold, X, thus raises the aggregate income of the economy and, at the same time, equalizes the distribution of income. It is in that fundamental sense that a more equal distribution of opportunities, here the distribution of wealth, enhances economic efficiency. If, as a new entrepreneur, P expands his or her business over time, then this equalizing of the distribution of wealth leads to faster economic growth.

An important point in the preceding argument is that it is not the inequality of the wealth distribution per se that is responsible for economic inefficiency, but the inequality in the access to credit. In the preceding example, a micro-credit organization lending money to P so that P undertakes his/her high-return investment project would lead to the same efficiency gain. This argument generalizes to other income-generation facilities: unequal access to education, gender or ethnic discrimination in the labor market, unequal access to cultivable land, to justice when seeking to protect one's property, to public decision making in case one is concerned with a specific public good, etc. In all these cases, income-enhancing opportunities are missed by part of the population, even though their private and social return could be higher than in the

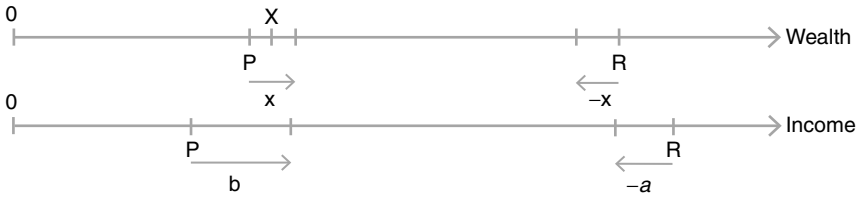


Figure 2.1 Wealth transfer and efficiency
Source: Author.

rest of the economy. Overall, the economy produces less and, possibly, grows more slowly than it could.

Another important consequence of the argument illustrated in Figure 2.1 is that the equalizing of opportunities, in that case through wealth redistribution, generates both a more efficient economy and a less unequal distribution of income. Therefore, it may be the case that a more inegalitarian distribution of income is associated with less efficiency, but the association is not causal. Both actually result from the inequality of opportunities. In the same way, discrimination on the labor market may prevent employing productively workers from a given gender or ethnic group. Again, this reduces the total output in the economy and creates income inequality.

Such are the fundamental justifications of the first message in the list above. I'll come to the second one in section 3 below. As for messages (iii) and (iv), they derive quite naturally from the first two messages. In presence of inequality of opportunities of the type mentioned above, policies for a better and faster development clearly involve the equalizing of opportunities or the leveling of the playing field in several key dimensions – from education and the credit market to the fight against discrimination, to an equitable distribution of infrastructure or to a well-functioning land market.

The word 'politically' in message (iii) is important. It seems to depart somewhat from the kind of argument that justifies the equalizing of opportunities as one key aspect of long-run development strategies. But, actually, it does not. It is directly related to the point made above about unequal access to public decision making being a possible source of economic inefficiency. This is the case because people unable to have their voice heard will not be able to convince the community that it may be more efficient to invest in a public good, say an infrastructure, that benefits them but at the same time would contribute to expanding total income. This is true at the local level as well as at the national level. It is unlikely that a country run undemocratically by a predatory elite will take decisions that will enhance economic growth for all. Decisions will be those which favor the elite, and not necessarily the whole population. This is the reason why the WDR 2006 placed so much emphasis on the issue of the quality of institutions and governance. But, actually, this must be considered as obeying the same logic as the argument in Figure 2.1: unequal access to facilities or decisions that influence the income generation of part of the population – in general, the poorest one but in some cases also the middle class – is a source of economic inefficiency and, in most cases, of income inequality.

On that line of argument, the WDR 2006 then examined the evidence in support of the general ideas above and explored the nature of the policies able to equalize opportunities, their potential and limitations, their likely effect on efficiency, growth and the income distribution but also their cost. The policy message was threefold: (i) tilting the accumulation of human capital or

infrastructure or the development of credit towards those groups in the population that had limited access to them; (ii) correcting market failures responsible for unequal opportunities; (iii) making institutions efficient and receptive of the interest of least favored citizens.

The message was simple and represented a rather natural inflexion of standard macro-oriented accumulation policies towards income and opportunity poor people. It was well received. At the same time, the message may have not insisted enough on several related issues, and in particular the issue of financing better opportunities in underprivileged groups and hence the role of taxation and income redistribution. Before getting to this, however, I will come back to the basic argument about the complementarity between equity and efficiency, first from a conceptual and then from an empirical point of view, as some of its implications may not have been stressed strongly enough in the report.[3]

2.2 The equity–efficiency complementarity: which equity, which efficiency?

Terminology is important in writings that try to introduce conceptual innovations. Without clear definitions, there is a risk of misunderstanding. Both equity and efficiency were given in the report a meaning which was not necessarily the most widely used among development economists and practitioners.

Equity throughout the report was to be understood mostly as the equality of opportunities. This concept initially was somewhat instrumental in the sense that it was explicitly proposed as an alternative to the more widely used concept of income equality. More fundamentally, however, equity was an ex ante equality concept defined in the space of 'opportunities' or 'capabilities' rather than in the space of 'results', as would be the case with incomes, consumption expenditures or any standard measure of economic wellbeing. Normatively, it was clearly related to the social justice literature: the difference between circumstances and efforts in Roemer (1998), the notion of freedom in Sen (1999) or the primary goods in Rawls (1971). To be closer to Rawls, however, equity in the WDR 2006 was defined not only by the equality of opportunities (or circumstances to use Roemer's terminology) but also by the 'avoidance of extreme deprivation in outcomes', so that the pursuit of this normative criterion was fully consistent with the objective of eradicating poverty.[4] Yet, it must be kept in mind that other authors would use the concept of equity in a different way, referring in particular to Sen's (1973) 'weak equity axiom' roughly requiring to give most to the most disadvantaged and leading to criteria close to Rawls' Maximin principle.[5]

On the efficiency side, the concept used throughout the report was clearly that of 'aggregate' efficiency. As illustrated in the previous section, equalizing opportunities, or increasing equity, was thus seen not only as ethically good

but also as instrumental in increasing aggregate resources in the economy, or, more bluntly, GDP. Yet, as very rightly pointed out by Roemer (2006), there may be some inconsistency in combining aggregate efficiency and equality of opportunities as simultaneous goals of development policies. On the production possibility frontier, two different goals cannot be pursued simultaneously, except in very particular circumstances. In general, if a society seeks to maximize equity, it must be willing to sacrifice some aggregate income. If it wants to maximize aggregate income, it must be willing to accept some inequity. Instead of considering that aggregate efficiency is something good per se, the social relevance of that complementarity between equity and efficiency, which was the cornerstone of the WDR 2006, has to be justified more rigorously.

The justification explicitly given in the report is that the economies under scrutiny are not on their production possibility frontier, precisely because of the inequality of opportunities and market failure as in the example of the previous section. Starting from within the production possibility domain, there thus is room to improve both equity and aggregate efficiency as commonly argued in the literature on 'efficient redistribution'.[6] Choosing to increase equity and efficiency rather than equity only possibly at the cost of aggregate efficiency is another matter. This may be justified in various ways which could have been discussed in more depth in the WDR 2006.

The first justification would be to simply rely on the well-known social welfare function proposed by Sen where the mean income of society is simply weighted by a decreasing function of income inequality. The only thing would be that the inequality of income would be replaced by the inequality of opportunities. Formally, Sen's social welfare function can be written as:

$$W = y.(1 - G_y) \tag{1}$$

where y stands for the mean income of the population and G_y for the Gini coefficient of the income distribution. By contrast, the social welfare function consistent with the main argument in the WDR 2006 would write:

$$W^* = y.(A - I_\omega) \tag{2}$$

where I_ω stands for the inequality of opportunities, assuming it can indeed be represented through a scalar measure[7] with maximum value A. In that formula, aggregate efficiency, y, and equity $(A - I_\omega)$ are clearly substitutes. But, one must take into account that the inefficient functioning of the economy may imply, at least locally, a negative relationship between aggregate efficiency, y, and the inequality of opportunity, I_ω, as in the argument of Figure 2.1. In that case, maximizing social welfare clearly implies simultaneously reducing the inequality of opportunity and increasing aggregate efficiency, at least as long as the

negative relationship between γ and I_ω holds. The complementarity between equity and aggregate efficiency is actually imposed by the positive side of the economy, not the social valuation criterion.

An alternative justification of the equity–efficiency complementarity, not very far from what is actually observed in the world, would be to consider that aggregate efficiency or strictly positive GDP growth are political constraints that bear on equity-enhancing policies. This is certainly not unrealistic, given the practical importance of the GDP rate of growth to assess the performance of a government, from within and without a country. Within a country, however, this raises deeper questions about the political economy motives behind such a constraint. An equity-enhancing development strategy would clearly be opposed if it were harming the elite in power. The rate of growth entailed by such a policy must be sufficiently high so as to compensate, and preferably over-compensate the elite.

Finally, efficiency may be interpreted precisely in the Paretian sense that is usual in economics. According to that criterion, no reform could be considered as a social improvement if it harms at least one economic agent. In the example of Figure 2.1, the wealth redistribution is not Pareto efficient because individual R is worse off, even though aggregate efficiency increases, i.e. $b-a>0$. But, precisely because the reform is efficient at the aggregate level there is room for compensating the loser R and keeping some benefit to the winner, P. Thus, aggregate efficiency is *potentially* Pareto efficient.

Some authors were satisfied with this argument and were willing to declare that aggregate efficiency was a good social valuation criterion, even though the actual compensation of the loser by the winner would not necessarily take place. This is the well-known Kaldor–Hicks efficiency or compensation criterion. If one does not want to go that far, however, and having in mind the political economy of the reform that implies that R would oppose it if he or she were not compensated, then it is necessary to consider the way the compensation is made.

The report does not really consider this issue of the compensation, the channels that could be used and the likely distortion costs involved. On the contrary, it insists that equity-enhancing policies considered in the report may lead to aggregate efficiency gains but not to a Pareto-superior situation. These policies will thus be opposed by those who have to lose from them, for instance those who are financing them, like R in Figure 2.1. Such policies would thus be undertaken only with a government following a social objective of type (2) above and able to impose its view on losers. This is indeed why the issue of institutions and governance is given so much importance in the report.

To make things clearer, let's introduce the distortion costs arising because of the practical impossibility of lump-sum transfers in real economies. It is now assumed that the transfer of x from R to P involves a distortion cost equal to

an annualized income loss c, so that the actual net total gain in annualized incomes is now equal to $b-a-c$, which is still assumed to be positive. Now, in order for R to be compensated for its loss, P must transfer a back to R, which entails an additional inefficiency cost equal to d. It is now quite possible that the remaining surplus $b-a-c-d$ is not positive anymore. Then, given that lump sum transfers generally are not feasible and transfers involve efficiency losses, the initial aggregate efficiency gain cannot be transformed into a Pareto improvement. For the final surplus to be positive, it may be necessary that the transfer channels be sufficiently efficient and also that the government credibly commit to activate the a transfer from P to R when the initial surplus $b-a-c$ will materialize. Both conditions actually require the existence of satisfactory redistribution institutions, which is questionable in many low-income countries.

Yet there is an interesting and extremely relevant instance where the a transfer from P to R may not be necessary to produce a Pareto improvement. It is when the project undertaken by P after the wealth transfer x involves externalities that benefit R. In the case of education, for instance, it may be the case that R is a capital owner who will benefit in the long run from a more educated labor force that will increase the rate of profit. If this offsets the initial cost, a, then R may be better off at the end of the process. R should thus accept to transfer x to P so that P's children to go to secondary school or college, except of course if there is very much uncertainty on the size of the future externality[8] or if R has a time discount rate very different from the rest of society.

Somehow, it is something of this type that was implicitly referred to in the report when arguing that equity-enhancing policies could have a positive impact on the rate of growth of the economy and this acceleration would benefit everyone in the population, including those who will be initially losing in launching these policies. Somehow, however, the report did not go into the distributional features of the additional growth that would be triggered by the reduction in the inequality of opportunities. It rightly argued that there might be a time lag between the short-run distortion cost generated by an opportunity-equalizing policy and its potential long-run efficiency benefits and the recommendation was thus to make sure that these long-run efficiency benefits would be properly taken into account when evaluating the economic impact of equity-enhancing policies. However, the issue of the distribution of these long-run benefits was not really discussed. The same is true for the possible heterogeneity in time discount rates that may sometimes be too high for these benefits to compensate the short-run cost of the policy among losers.

In summary, unlike what was suggested by Roemer in his review of the WDR 2006, there was no real conflict between equity and aggregate efficiency in the implicit social welfare function in the report. The recommendation to pursue both objectives was consistent with the explicit assumption that most

economies find themselves inside the production possibility frontier and the fact that some equalizing of opportunities could get it closer to the frontier. In other words, there is ample room for improvement in both the equity and the GDP space.

Even so, equity-enhancing policies that also increase aggregate efficiency involve winners and losers. Compensating the losers may be costly and adequate redistribution channels may not be available. Implementing such policies thus requires a government able to follow a social welfare objective à la Sen, as in (2) above, which, in turn, requires the necessary institutions and governance for this not to be opposed by a losing elite. A better outcome is when the policies that equalize opportunities create externalities that benefit the whole society, as possibly with education. Such policies are potentially Pareto superior, except for the fact that it might take time before the short-term losers are fully compensated. A more detailed distributional analysis would be necessary to investigate such possibility. This was not attempted in the report in view of its obvious difficulty and country specificity.

2.3 Evidence about the relationship between inequality, efficiency and growth

This reference to the possible growth externalities of equity-enhancing policies logically leads to the issue of the nature of the evidence about the efficiency gain of these policies.

This evidence was extensively discussed in the WDR 2006. A host of examples were given where correcting market failures led to progress in both equity and aggregate efficiency, very much along the lines of the imperfect credit market argument behind Figure 2.1 above: credit market imperfections leading to observed rates of return to capital higher in small informal firms, lower schooling enrollment or lower school quality among poor people slowing down human capital accumulation, share-cropping contracts dis-incentivizing production, caste, gender or ethnic discrimination discouraging efforts, and so on.

The problem with this evidence is that it is essentially microeconomic in nature and says nothing about the overall effects such policies could have on the level of GDP or its growth rate if they were undertaken at the national level. Yet such information is necessary to compare these policies to other policies that may not improve equity directly but may have a much larger effect in terms of GDP and possibly of poverty. How does a policy subsidizing credit to small and medium-sized enterprises (SMEs) compare with public infrastructure investments, or a market land reform compare with a major irrigation program?

The point here is not to get back to the issue of the social welfare function and the relative weights of equity and efficiency within it. The point is to

stress that those policies with both an equity focus and a positive effect on efficiency cannot be compared to policies with a pure efficiency focus because little is known about the aggregate output effect of the former. We have some idea about the GDP impact of an infrastructure investment program, through a simple production function approach or cross-country growth regressions. For instance, Calderon et al. (2015) found that the elasticity of GDP to an infrastructure capital indicator in a panel of countries is .08 while in an earlier paper Calderon and Serven (2004) found that infrastructure capital also reduced income inequality, suggesting that this type of investment, which involves high public sector participation, indeed contribute simultaneously to equity and efficiency.[9] Among equity-enhancing policies, comparable estimates are available for education focused policies – see in particular Cohen and Soto (2007) who reconcile micro- and macro-based estimates of the return to education – or health – see, for instance, Weil (2007). But, what do we know about the aggregate impact of those policies directed toward social protection, SME development, agricultural extension, market land reform, anti-discrimination or local empowering through decentralization?

Symmetrically, it must be recognized that the knowledge of the distributional impact of policies more directly oriented towards aggregate growth like trade, industrial or infrastructure policies is often limited. Both equity- and growth-focused development policies call for more work to know better their joint macro- and micro impact.

The particular case of human capital policies needs to be stressed as they often tend to combine both the equity and the growth objectives and may make them almost inseparable in some cases. In many countries, progress in the overall school enrollment rate is likely to concern mostly the underprivileged. This is simply because of the (inequitable) political economy phenomenon by which the progressive expansion of a public service like public education or health care sequentially benefits social groups with a declining political influence. In effect, it would often be difficult to say whether human capital policies are equity oriented – i.e. benefiting unfavored groups, or essentially growth-oriented.

Apart from the preceding examples, the difficulty of identifying the growth effect of equity-enhancing policies[10] makes it tempting to rely on the voluminous literature about the relationship between income inequality and economic growth as revealed by cross-country comparisons. After all, as the inequality of opportunities was seen as a major determinant of the inequality of incomes, the latter might taken as being be a good marker of the former. Unfortunately, even if this were the case, which is not granted, this would not be very much help.

The existing evidence is somewhat ambiguous with some papers pointing to a negative and others to a positive relationship depending on whether the

analysis bears on pure cross-sections or panel data, and also on the data sets being used. Today, it would seem that an agreement is building up on the existence of a long-term statistical negative relationship. The most recent contribution to that literature – the chapter by Ostry in the companion volume – based on a large sample of countries and covering an extensive time period indeed confirms such a relationship, although the database they use is not without problems.[11] However, the problem is that, even though this relationship is in agreement with the theoretical argument development above, its policy implications are limited. In addition to the usual worry about cross-country regressions that ignore the country heterogeneity in the relationship being tested, the reason is that we are in front of a reduced form model which gives no information on the policies able to modify the distribution of income and the pace of economic growth. These may be the equity-enhancing policies discussed in the WDR 2006 or those policies able to directly modify the distribution of income, as indeed argued by Ostry et al. (2015) on the basis of the observation that income redistribution does not seem to affect economic growth once the effect of income inequality after redistribution has been accounted for. Yet, I argued in my comments on their paper, in this volume too, that this argument was not fully convincing. More structural models with explicit policy instruments are needed to understand in more depth the nature and the implication of the statistical relationship that seems to exist between income inequality and economic growth.

2.4 Equalizing opportunities and income redistribution

The emphasis put by the WDR 2006 on the inequality of opportunities led to relatively little space devoted to income inequality and income redistribution as such. To be sure, the report insisted on the need to rely on taxation to raise the money required by additional public spending directed towards equalizing opportunities. But, relatively little space was overall devoted to the issue of the progressivity of the taxation system and public expenditures. Also, the fact that equity, as defined in the report included a no-material deprivation dimension – to prevent low future opportunities of a family and certainly of the children – makes income redistribution an indispensable equity-enhancing instrument.

Unlike income, opportunities cannot be easily redistributed. Going back to the example of Figure 2.1, the initial redistribution of wealth from R to P is not an option except at a huge social cost and putting the private property principle and the economic system built upon it at serious risk. What is possible, however, is to tax the income of R and to subsidize credit to P, in effect taking *a* away from R's income and giving it to a financial intermediary which will then accept to lend money to P. In other words, equalizing opportunities may

go through taxation and public spending. As a matter of fact, many equity-enhancing policies thus have a dimension of income redistribution.

This is obviously the case for education and health. In these areas, equalizing opportunities goes through public spending on the accumulation of human capital among disadvantaged groups, which requires raising funds through taxation of income or assets. As just mentioned, the same can be said of the redistribution of assets like financial wealth or land. Social protection and anti-poverty cash transfers are other areas where preventing the inequality of opportunities to worsen through poverty traps requires income transfers from better off households. Yet other dimensions of the inequality of opportunities may not necessarily require such transfers. Fighting discrimination and monopolistic positions on all types of markets asks for adequate governance more than anything else.

That equalizing opportunities often goes through taxation and public spending makes the potential for income redistribution through the fiscal system all the more important. The main point in the WDR 2006 that the equality of opportunities rather than that of income matters for economic growth and development is right, but promoting the equality of opportunities actually requires a substantial redistribution of income.

Evaluating the extent of equity-enhancing policies thus asks for a thorough analysis of the whole redistribution that takes place though taxation and that part of the public expenditures whose beneficiaries can be easily identified. It matters not only who is being taxed but which part of the population benefits relatively more from public spending and the change it makes in the present or future distribution of opportunities. In this respect, the WDR 2006 was right in emphasizing that in many countries the fiscal system was extremely regressive either on the taxation side, for instance when small farmers are being taxed through unfair prices imposed by marketing boards, or on the public spending side, in the multiple instances where public spending benefits well-off families which could easily afford the cost of the services freely supplied by the public sector.

A common recommendation on taxation is to have a broad base and low tax rates so as to maximize receipts and minimize distortions. This seems somewhat contradictory with what is learned from optimal taxation theory. Very general social welfare functions and full account being taken of the efficiency loss due to too high tax rates lead to progressive tax systems. There is no reason this would be different if welfare functions were cast in terms of average income and inequality of opportunities as in (2) rather than in terms of individual incomes. At the bottom of the income scale, any decrease in income would mean a loss of opportunities, by creating more poverty and reducing the investments in the human capital of children. At the top, it is less likely that access to income generating activities would be restricted by a tax rate higher than for the rest of the population, even though incentives might be weaker.

If the administrative capacity permits to manage a progressive income tax system, the pursuit of equity would command to indeed introduce such a tax. Modern payment technology have made the control of income and expenditures easier, but the political will to implement a progressive tax system based on income or assets, or to fight tax evasion and corruption, is often lacking.

Progressivity may also be implemented on the public spending side and in some cases this may be easier than on the tax side. There are numerous possibilities depending on the country being considered: recovering the cost of higher education for children from well-off families while granting scholarships to students coming from more modest households, substituting food or energy subsidies by targeted cash transfers, cutting on top pensions in case of a deficit in a pay as you go pensions system, and so on. All these measures free resources at the expense of the upper part of the income distribution, which can be mobilized to improve opportunities at the bottom of the distribution by spending on education, health, infrastructure, credit to SMEs, and so on. In all these cases, progressivity may be achieved without a complex income control apparatus. Scholarships or fee exemption in the higher education system, or replacing consumption subsidies by cash transfers, only requires identifying households in the lower part of the distribution. Such an identification system already exists in many countries, in particular in those which have implemented cash transfer systems. On the other hand, cutting on top public pensions is always possible by introducing a ceiling in the payments made by the social security system.

The WDR 2006 could have put more emphasis on this distributional aspect of the funding of equity-enhancing policies and the relationship between income redistribution through fiscal policy and equalizing opportunities. True, it already covered very much ground and this kind of incidence analysis of taxation and public spending in developing countries has already been applied in different contexts – see, for instance, Bourguignon and Pereira da Silva (2003). The point is, however, that the incidence analysis of fiscal policies is very often incomplete. Even though it was not very explicit on this issue, an important contribution of the WDR 2006 has been to inspire an extension of standard benefit incidence analysis to opportunity incidence analysis – see, for instance, Cuesta (2014) – and the design of the Human Opportunity Index described in Paes de Barros et al. (2009) and Molinas et al. (2010).

These extensions enlarge the analytical framework for the evaluation of development. When restricting development to its pure income dimension, it had been a progress to move from the GDP rate of growth as a single indicator of development progress to the 'growth incidence curves' that described the way in which income had increased at the various rungs of the income ladder.[12] It is an additional improvement to move beyond a strict income criterion and to include the change in the distribution of opportunities in evaluating development progress.

At the same time, the link between the distribution of income and the distribution of opportunities should be investigated in more depth. The causal relationship that goes from the distribution of opportunities to the distribution of income is pretty clear. The reverse relationship may be more complex because it involves several mechanisms. First, as mentioned above, low absolute or even relative incomes are synonymous of a lack of future opportunities for those in that part of the income distribution and their children, so that it is difficult not to include some measure of income inequality or poverty in the evaluation of opportunities. Second, when measuring the opportunities faced by children, as for instance within the Human Opportunity Index, the income per consumption unit in their family is considered as a 'circumstance' beyond their control, in the same way as gender or ethnicity or the area where they live. Somehow, the distribution of income is thus one of the determinants of the inequality of opportunities faced by children. Third, as already stressed, the various dimensions of opportunities often are dependent on public spending financed by taxation so that equalizing opportunities necessarily requires some income redistribution.

A last link between the inequality of opportunities and the distribution of income goes through what is called 'horizontal inequality' in income distribution analysis. It corresponds to differences in earnings between individuals with different circumstances, but the same achievements in outcome dimensions other than income. Typically, this would be the earnings differential between two workers with the same education, the same job experience and similar occupations but different gender, ethnic origin or even family background. It is true, however, that, unlike in the preceding case, the causality runs here mostly from the inequality of opportunities to that of income, unlike in the preceding case.

2.5 Top incomes, institutions and governance

Even though top incomes had not yet been brought to the forefront of the debate on inequality and growth or economic efficiency and were thus not considered explicitly in the WDR 2006, they really were like the elephant in the room! Although it was clearly central, the issue of the share of the top percentiles of the distribution of income was difficult to address in a report directed towards policy makers and politicians.

The embarrassment does not arise so much from having in some cases to tell governments that the economic elite is too rich and too corrupt and should be reined in more strongly and taxed more heavily – assuming, of course, that there is an unequivocal proof that it is because of their excessive income share that growth and poverty reduction are too slow. The difficulty is simply that it seems counter-productive, and at best useless, to recommend to policy makers a set of

measures that they cannot enforce because of the political economy constraints under which they operate. This is the familiar difficulty of political economy analysis: what kind of policy recommendation can you draw from an analysis where policy decisions are essentially endogenous? In the present case, how is it possible to recommend reforms leading to a broader sharing of political power where the latter is precisely monopolized in the interest of a small elite?

Even though the WDR 2006 did not deliver such a message directly, it was very explicit in dealing with institutions and governance on the danger of the "concentration of political power in the hands of a narrow group or an elite". This was seen as the cause of bad institutions that would encourage the creation of rents through all kinds of monopoly situations and the disregard of all claims made by the poor or the middle class, except perhaps in the presence of threats on the power of the elite. The consequence of such a situation could only be an overly unequal distribution of income and an inefficient and slow growing economy.

The argument in the WDR 2006 about the deleterious economic effect of too much concentration of political power and bad institutions was cast in terms of equity. The link with the inequality of income thus seemed to go in a single direction. The reverse causation between the concentration of income and the concentration of political power through the income elite-influencing politicians received less explicit attention. Yet, if such a link does exist, then the logic of the WDR 2006 would have to be reversed for a while: income inequality would be causing economic inefficiency and inequality in the distribution of political power rather than the opposite.

Is there evidence of such a link that goes from income concentration to political power concentration? This is difficult to ascertain because of the obvious two-way causality between these two phenomena and the vicious circle they create. This is the vicious circle discussed in Acemoglu and Robinson (2006) by which weak political institutions generate slow growth and a high concentration of income, which leads, in turn, to weak institutions and bad governance. But how does a country gets into the vicious circle is difficult to say. Engerman and Sokoloff (2000) have held that weak institutions in Latin America were the consequence of a high initial level of inequality at the time of the 'conquista' and were then responsible for the persistence of inequality.[13] Other authors would argue that the concentration of political power was the root cause of the vicious circle, as inequality may not have been as high in the early days of independence but increased very much afterwards.[14] More generally, it is not clear whether there are examples of countries where some exogenous non-political shock was responsible for a substantial increase in inequality and led after some time to explicit signs of political power concentration.

An interesting attempt has been made by Chong and Gradstein (2007) to disentangle the causality relationships between income inequality and the quality of institutions using time series analysis and Granger causality techniques on a

panel of developing countries. Not surprisingly, they found that causality was going in both directions. Interestingly enough, however, it was much stronger in the direction of income inequality causing weak institutions than the other way around. Unfortunately, this kind of cross-country analysis is never fully convincing, especially in the present case where causality tests are based on the time series properties of variables with typically little time variations. In addition, the imprecision of the data on both the various dimensions of institutional quality and income inequality is most likely to bias regression estimates.

In this respect, it is worth stressing that the true level of inequality in most developing countries is likely to be severely underestimated because top incomes are under-represented and/or under-reported in standard household surveys. This bias is also present in developed countries but to a lesser extent and corrections are possible using tax return statistics. Because a large part of the missing income comes from capital and thus tends typically to fluctuate more than labor income, recorded statistics probably miss important time variations in income inequality.

The preceding results on the relationship between income inequality and institutions refer to the quality of institutions as summarized by the various indicators published in the Worldwide Governance Indicator database[15] – voice and accountability, government effectiveness, rule of law, corruption, … – rather than to the nature of political institutions, and in particular the extent of democracy. Attempts at establishing a relationship between inequality and democracy have generally failed, one of the reasons being that, as argued by Acemoglu et al. (2014), a democratic political system is compatible with different levels of concentration of de facto political power.

Income inequality has also often been mentioned as the possible cause of social tensions even in well-governed countries. After all, the "we are the 99%" slogan of the Occupy Wall Street crowd was a protest against the concentration of income, and possibly of economic, if not political power, in the hands of the top 1%. Apparently, this movement was not successful in launching significant structural reforms, but things might be quite different if inequality keeps rising in the United States and sparks new protests in the future. It is also possible that opinions and the will for collective action evolve at a slow pace. For instance, the unexpected public success in the United States and elsewhere, of such a scholarly book on inequality as Piketty's *Capital in the 21st Century* may well be reflecting the rising consciousness of the importance of maintaining inequality at 'acceptable' levels in our societies, in the line of the Occupy movement. In Spain, the rise of the 'Podemos' party born from the 'indignados' movement, contemporary of Occupy, shows that such apparently marginal protests may reveal much deeper concerns in the public opinion.

It must be recognized that we know little about how grassroots and middle-class movements do, or do not, influence social and economic policy decision

making on inequality issues. There presumably is a level of the inequality of income, possibly as the sign of inequality in opportunity dimensions, and in some specific opportunity dimension - that will be found intolerable by society, so that social tensions will appear when inequality will get close to that threshold. The problem is that this threshold is essentially unknown.

2.6 Conclusion

The WDR 2006 has been instrumental in bringing inequality issues to the centre of the development policy debate with the important caveat that not all types of inequality were equivalent. Its main contribution is indeed to have shown the importance of reducing the inequality of opportunities, as measured by the inequality in the distribution of economic endowments in the population and in their access to various income-generating facilities, from both an intrinsic and an instrumental point of view. At a point of time, it is the distribution of opportunities that determines both the aggregate efficiency of the economy and the distribution of current income. Accelerating development may thus be obtained by reducing the inequality of opportunities and the report devoted considerable space about policy measures to achieve such a goal.

Such a strong message may have had the effect of diverting the spotlight from the other dimension of inequality, that of outcomes and, in particular, income or standard of living. This paper has tried to redirect the spotlight on the dual aspect of inequality, insisting upon the need to look simultaneously at both the inequality of opportunity and the inequality of outcomes, or, more generally, at the whole vector of inequalities, whether they are concerned with income, wealth, gender or ethnic discrimination, exclusion from some markets or the inability to influence public policy making, including the functioning and the role of the state. Such a comprehensive approach to inequalities is needed for policy making as most attempts at equalizing any one of the dimensions of this vector of inequalities will require some change in other dimensions and, in particular, the income dimension. Vice versa, any change in income inequality will require changes in several other inequality dimensions. The same generalized approach to inequality is to be adopted by researchers in the field of economic development, even though the relationship between these various types of inequality is certainly much more complex and difficult to observe and analyze than inequality in a single outcome or opportunity dimension, as presently studied.

It is quite encouraging to see that, following the WDR 2006, a substantial progress has been made in measuring and analyzing the inequality of opportunity and that this topic is much more present in the development policy debate than this was the case before. Yet considerable work remains to be done for this area of research to acquire the same maturity as the field of income inequality.

As a matter of fact, the main conclusion of this chapter is probably that the most important challenge to be addressed at this stage is the integration of the various aspects of inequality within a unified and encompassing view.

Notes

1. See, for instance, Ravallion (2001; Ravallion and Chen, 2003) or Bourguignon (2003) – see also the 'poverty-inequality-growth triangle' in Bourguignon (2004).
2. I am especially grateful of course to Francisco Ferreira and Michael Walton who headed the team.
3. A more technical discussion of somewhat different questions opened up by the report can also be found in Bourguignon et al. (2007).
4. The avoidance of extreme deprivation may also be considered as a requirement for maintaining the opportunity for all to simply function.
5. See, for instance, Hammond (1976). Bourguignon et al. (2007) showed how the equity criterion in the WDR 2006 can be actually be interpreted in terms of the Rawlsian Maximin.
6. See, for instance, Bardhan, Bowles and Gintis (2000). The full argument may be found in the report in the technical box in p. 78.
7. This is of course quite far from certain. Ferreira and Gignoux (2011) propose an estimate of a lower bound of the opportunity share of income inequality.
8. Note that, as stated, the argument does not refer to an externality but to the general equilibrium effect of a larger educated labor force on the rate of return to capital. In that case, however, workers who initially had gone to college will witness a drop in their wage rate. An educational externality à la Lucas (1988), i.e. which increases the total factor productivity, is necessary to generate the Pareto improvement discussed here.
9. Note, however, that the econometric estimation methodology in the 2004 paper was much less advanced than in the more recent paper by Calderon et al. (2014), which, unfortunately, does not consider the effect of infrastructure on inequality for lack of annual series on inequality.
10. Marrero and Rodriguez (2013) and Ferreira et al. (2014) are two recent papers that try to estimate the effect of some measure of the inequality of opportunity on growth, across the US states for the former, across countries for the latter. In both cases, the problem is of course that only a lower bound of the inequality of opportunities is known. A negative and significant effect is found in the former study while the results of the latter prove more elusive.
11. See my review of their paper in this volume.
12. See Ravallion and Chen (2003).
13. Easterly (2007) also suggests that there is some econometric evidence suggesting that the causality runs from income inequality to bad institutions.
14. The view that institutions were the entry point into the vicious circle is illustrated by Acemoglu et al. (2005). On Latin America, Williamson (2009) holds that inequality was much lower at the time of independence in Brazil or Peru than it is today. Yet, it seemed to be higher in Mexico and Chile.
15. See the methodology for the production of these indicators in Kaufmann et al. (2002).

References

Acemoglu, Daron, Simon Johnson, and James A. Robinson (2005) "Institutions as a Fundamental Cause of Long-Run Growth," in Philippe Aghion and Steven Durlauf (eds), *Handbook of Economic Growth, Vol. 1A*. Amsterdam: Elsevier, pp. 385–472.

Acemoglu, Daron and James Robinson (2006) *Economic Origins of Dictatorship and Democracy*. Cambridge: Cambridge University Press.

Acemoglu, Daron, Suresh Naidu, Pascual Restrepo, and James A. Robinson (2014) "Democracy, Redistribution, and Inequality," in A. Atkinson and F. Bourguignon (eds), *Handbook of Income Distribution Vol. 2B*. Amsterdam: Elsevier, pp. 1869–949.

Aghion, Philippe and Patrick Bolton (1996) "A Theory of Trickle-Down Growth and Development," *The Review of Economic Studies*, vol. 64, no. 2, pp. 151–72.

Aghion, Philippe, Eve Caroli and Cecilia Garcia-Penalosa (1999) "Inequality and Economic Growth: the Perspective of the New Growth Theories," *Journal of Economic Literature*, vol. 37, no. 4, p. 1617–60.

Bardhan, Pranab, Samuel Bowles and Herbert Gintis (2000) "Wealth Inequality, Wealth Constraints and Economic Performance," in A.B. Atkinson and F. Bourguignon (eds), *Handbook of Income Distribution, Volume 1*. Amsterdam: Elsevier, pp. 541–603.

Bourguignon, François (2003) "The Growth Elasticity of Poverty Reduction: Explaining Heterogeneities Across Countries and Time Periods," in T. Eicher and S. Turnovsky (eds), *Inequality and Growth: Theory and Policy Implications*. Cambridge, MA: MIT Press, pp. 3–26.

Bourguignon, François and Luis Pereira da Silva (2003) *The Impact of Economic Policies on Poverty and Income Distribution: Evaluation Techniques and Tools*. Washington, DC: World Bank and New York: Oxford University Press.

Bourguignon, François (2004) *The Poverty–Inequality–Growth Triangle*, Working Paper No. 125, New Delhi: Indian Council for Research on International Economic Relations (ICRIER).

Bourguignon, François, Francisco Ferreira and Michael Walton (2007) "Equity, Efficiency and Inequality Traps: A Research Agenda," *Journal of Economic Inequality*, vol. 5, pp. 235–56.

Calderon, Cesar and Luis Serven (2004) "The Effects of Infrastructure Development on Growth and Income Distribution," The World Bank, Washington, http://elibrary.worldbank.org/doi/pdf/10.1596/1813-9450-3400.

Calderón, Cesar, Enrique Moral-Benito and Luis Serven (2015) "Is Infrastructure Capital Productive? A Dynamic Heterogeneous Approach," *Journal of Applied Econometrics*, vol. 30, no. 2, pp. 177–98.

Chong, Alberto and Mark Gradstein (2007) "Inequality and Institutions," *The Review of Economics and Statistics*, vol. 9, no. 3, pp. 454–65.

Cohen, Daniel and Marcelo Soto (2007) "Growth and Human Capital: Good Data, Good Results," *Journal of Economic Growth*, vol. 12, pp. 51–76.

Cuesta, Jose (2014) "Social Spending, Distribution, and Equality of Opportunities: The Opportunity Incidence Analysis," *World Development*, vol. 62(C), pp. 106–24.

Dollar, David and Aart Kraay (2002) "Growth is Good for the Poor," *Journal of Economic Growth,* vol. 7, no. 3, pp. 195–225.

Easterly, William (2007) "Inequality Does Cause Underdevelopment: Insights from a New Instrument," *Journal of Development Economics*, vol. 84, pp. 755–76.

Ferreira, Francisco and Jérémie Gignoux (2011) "The Measurement of the Inequality of Opportunity: Theory and Application to Latin America," *The Review of Income and Wealth*, vol. 57, no. 4, pp. 622–57.

Ferreira, Francisco, Christoph Lakner, Maria Ana Lugo, and Berk Özler (2014) *Inequality of Opportunity and Economic Growth: A Cross-Country Analysis*, IZA, Discussion Paper No. 8243.

Galor, Oded and Joseph Zeira (1993) "Income Distribution and Macroeconomics," *The Review of Economic Studies*, vol. 60, no. 1, pp. 35–52.

Hammond Peter J. (1976) "Equity, Arrow's Conditions, and Rawls' Difference Principle," *Econometrica*, vol. 44, no. 4, pp. 793–804.

Kaufmann, Daniel, Aart Kraay and Massimo Mastruzzi (2002) *The Worldwide Governance Indicators: Methodology and Analytical Issues*, Policy Research Working Paper N° 5430, The World Bank.

Lucas, Robert (1988) "On the Mechanics of Economic Development," *Journal of Monetary Economics*, vol. 22, pp. 3–42.

Lustig, Nora and Sean Higgins (2013) *Commitment to Equity Assessment (CEQ): Estimating the Incidence of Social Spending, Subsidies and Taxes. Handbook*, CEQ Working Paper No. 1.

Marrero, Gustavo and Juan Gabriel Rodriguez (2013) "Inequality of Opportunity and Growth," *Journal of Development Economics*, vol. 104, pp. 107–22.

Molinas, Jose, Ricardo Paes de Barros, Jaime Saavedra, and Marcelo Giugale (2010) *Do Our Children Have a Chance? The 2010 Human Opportunity Report for Latin America and the Caribbean*. Washington, DC: World Bank.

Paes de Barros, Ricardo, Francisco Ferreira, Jose Molinas, and Jaime Saavedra. (2009) *Measuring Inequality of Opportunities in Latin America and the Caribbean*. Washington, DC: World Bank.

Piketty, Thomas (2014) *Capital in the Twenty-first Century*. Boston, MA: Harvard University Press

Ravallion, Martin (2001) "Growth, Inequality and Poverty: Looking Beyond Averages," *World Development*, vol. 29, no. 11, pp. 1803–15.

Ravallion, Martin and Shaohua Chen (2003) "Measuring Prop-poor Growth," *Economic Letters*, vol. 78, no. 1, pp. 93–9.

Rawls, John (1971) *A Theory of Justice*. Cambridge, MA: Harvard University Press.

Roemer, John E. (1998) *Equality of Opportunity*. Cambridge, MA: Harvard University Press.

Roemer, John E. (2006) Review Essay, "The 2006 World Development Report: Equity and Development," *Journal of Economic Inequality*, vol. 4, pp. 233–44.

Sen, Amartya (1973) *On Economic Inequality*. Oxford: Clarendon Press.

Sen, Amartya (1999) *Development as Freedom*. New York: Oxford University Press.

Sokoloff, Kenneth and Stanley Engerman (2000) "Institutions, Factor Endowments, and Paths of Development in the New World," *Journal of Economic Perspectives,* vol. 14, no. 3, pp. 217–32.

Straub, Stéphane (2008) *Infrastructure and Growth in Developing Countries: Recent Advancesand Current Challenges*, Policy Research Working Paper No. 4460, The World Bank, Washington, DC.

Stiglitz, Joseph (2012) *The Price of Inequality: How Today's Divided Society Endangers Our Future*. New York: W.W. Norton and Company.

Weil, David (2007) "Accounting for the Effect of Health on Economic Growth," *The Quarterly Journal of Economics*, vol. 122, no. 3, pp. 1265–306.

Williamson, John (2009) *History without Evidence: Latin American Inequality Since 1491*, NBER Working Paper No. 14766.

Equity and Development: Revisiting the 2006 *World Development Report*

Martin Ravallion
Department of Economics, Georgetown University

In the ten years since the World Bank's 2006 World Development Report (WDR) *Equity and Development* was being written, inequality has become even more prominent in policy debates. This comment focuses on three questions:

- Was the WDR's emphasis on inequality of opportunity justified and have defensible approaches to operationalizing the concept emerged since then?
- What other aspects of inequality have emerged as relevant to policy debates?
- How does the WDR's perspective relate to the spreading concerns about rising high-end inequality?

The WDR's emphasis on inequality of opportunity

"Equity" is a big word. As the WDR recognized, we need to be more specific, especially when we carry that word into policymaking. The WDR built nicely on a strand of thought from philosophy and economics that offers hope for consensus around the need to redress inequality of opportunity (INOPP). This was an important contribution to development policy discussions. Roemer (1998) had argued that we need only worry about inequalities that stem from *circumstances* beyond an individual's control – those things that are not traceable to the individual's own choices. By this view, inequality of outcomes is to be tolerated as long as it reflects *personal efforts*.

There is a continuing debate on this view. Can we ever ignore inequality due to differing effort? People make mistakes. It is surely unimaginable that any civilized society would do nothing about extreme, possibly life-threatening, deprivations on the grounds that they are traceable to some mistaken choices by the persons concerned. Inequalities stemming from choices can hardly be banned from public redress.

The WDR largely shied away from such issues, and focused more on the instrumental case for equity – its relevance to other objectives. This perspective

was clearly seen by the authors as having greater traction within the mainstream development community. Of course, as soon as one frames the issue in instrumental terms, "equity" loses its place at the high table of development goals. But many of the things that are important to development are not clearly ends in themselves, including economic growth.

The main goal for which equity was seen as instrumentally important was economic growth. For some time the Bank had seen growth as important for poverty reduction, which had been seen as the overarching goal since the early 1990s. The WDR saw inequality as costly to progress against poverty. However, it did not see equity as solely of concern for poverty reduction.

The WDR took it to be self-evident that INOPP was the relevant dimension of inequality for making its instrumental case. This has never been entirely clear. Inequality of outcomes also matters to growth. For example, the way credit market failures impact on growth is via the initial distribution of wealth. And it is not just inequality that matters; market failures can entail that poverty impedes growth and this is consistent with cross-country evidence (Ravallion, 2012).

The evidence for the WDR's position on this topic was weak at the time, as noted by Bourguignon (2016). In the only paper since then to test the assumption made by the WDR, Ferreira et al. (2014) study economic growth in a large cross-section of countries. They confirm past evidence that income inequality is harmful to growth but they do not find that the component of inequality that they attribute to unequal opportunities is especially harmful.[1]

From a time soon after the WDR 2006 appeared, the nub of the matter was how one might go about operationalizing a Roemer-influenced measure of INOPP. In practice, advocates of this approach ignore "effort" as an explicit factor and focus instead on reduced-form regressions for outcomes – typically household income or consumption – on selected measures of circumstances. The predicted values are then used for measurement. The Bank's *Human Opportunities Index* is an example, with origins in the work of Paes de Barros et al. (2009). We are seeing more applications of this approach.

There are some continuing concerns related to both what we mean by "circumstances" and the role played by "effort." In practice, the reduced-form interpretation of the regression of outcomes on circumstances allowed the possibility that latent effort was a function of circumstances. We then forgo knowing if the estimated effect of circumstances is direct or via effort. We do not need to know that in order to proceed in measuring INOPP.

But have we reliably identified an effect of circumstances? The things we observe in data are likely to be an incomplete account of the relevant circumstances. It is often argued that the regression is likely to underestimate the contribution of circumstances, although this is not obvious. What is being identified as the effect of a "circumstance" may only be statistically salient to outcomes because it is correlated with latent effort. When the "circumstances"

are measured by a variable like parental schooling it may seem beyond doubt that it is exogenous to the incomes of the children when they become adults. But what if poorly educated parents install values in their children that discourage work effort? While some of us might dismiss that possibility on a priori grounds, others do not, and credible measurement of the contribution of circumstances must be reasonably robust to such counter-arguments.

The standard regression method used in this literature for measuring INOPP may well *overestimate* the impact of circumstances on outcomes, given that latent effort interacts with circumstances to determine outcomes.[2] We really can't say with much confidence that this approach has truly isolated the impact of circumstances or the share of the variance in outcomes that is attributable to the observed circumstances.

While econometric subtleties such as this rarely come into the discussion of INOPP, most people are aware of the concern about latent aspects of effort. In practice, the logic of opportunity egalitarianism can slip easily into blaming poor people for their poverty, and excusing rich people for their success on the grounds of some claimed effort they have made but which is hidden from our view.

There is also a nagging concern that the predicted value of income based only on circumstances is not in fact a valid metric of welfare if we take seriously the idea that people choose their effort (Ravallion, 2015b). And that idea is already embodied in the formulations of the reduced form approach to measuring INOPP. It seems odd to say that (on the one hand) people derive a disutility from effort, which they weigh against the income gains, yet (on the other hand) not to measure income consistently with the choice of effort, which requires an equivalent income formulation, as I have illustrated elsewhere (Ravallion, 2015b).

Even putting these concerns aside, what does the WDR's INOPP emphasis really mean for policy? The WDR did not, in my view, provide a compelling case (on moral or other grounds) for thinking of equity as an independent goal to poverty reduction as long as the concept of "poverty" was sufficiently broad, including reflecting inequality amongst poor people.[3] We are probably not going to recommend curtailing the opportunities of those with ample opportunities (as pointed out by Roemer, 2013). So our focus will naturally be on raising the opportunities for poor people. Thus we end up back with poverty reduction (in some reasonably broad formulation) as the overarching goal. We have come full circle.

Aspects of inequality that received less attention in the WDR

While INOPP provided a key conceptual foundation for the WDR's instrumental case against inequality, the bulk of the WDR's empirical measures were

about inequality of outcomes, specifically inequality of household income or consumption per person. Here the WDR largely followed the long-standing practice of relying on *relative inequality* measures, defined in terms of ratios of incomes or consumptions.

There are continuing concerns about the relevance of the types of measures found in WDR 2006 and elsewhere (including the *World Development Indicators*; see, for example, World Bank, 2013) to ongoing debates about equity and development.[4] By contrast to relative inequality, *absolute inequality* depends on the absolute differences in levels of living. It is not that one concept is "right" and one "wrong." They simply reflect different value judgments, as embodied in the axioms of inequality measurement.

In this light, let us consider the long-standing debates on growth and equity. A large body of evidence has accumulated indicating that economic growth in developing countries tends to be distribution-neutral on average, meaning that changes in relative inequality are roughly uncorrelated with growth rates.[5] Finding that the share of income going to the poor does not change on average with growth does not mean that "growth raises the incomes (of the poor) by about as much as it raises the incomes of everybody else," as claimed by the *Economist* magazine (May 27, 2000, p. 94). Given existing inequality, the absolute income gains to the rich from inequality-neutral growth will, of course, be greater than the gains to the poor. The empirical finding in the literature that growth tends to be inequality-neutral within developing countries will carry little weight for those concerned about absolute inequality, who prefer translation invariance to scale invariance. One expects an absolute measure to rise with growth, and fall with contraction. I confirm this using country-level data in Ravallion (2014b).

Past and ongoing debates about the distribution of the gains from growth appear to rest in no small measure on this (rarely discussed) conceptual difference in how inequality is defined. When citizens see a growth process whereby the rich guy can buy a new car and the poor guy can only fix his bicycle they may be justified in calling it "inequitable" even if relative inequality is unchanged. Recognizing this point may not resolve the debate, but it would make clearer what is at issue.

Unlike those who see inequality as relative, those who view it in absolute terms will expect to see a trade-off between reducing inequality and reducing poverty. Some growth-promoting *and* poverty-reducing policy reforms may well come in for serious criticism from, and even be blocked by, a non-negligible number of observers concerned about widening gaps in living standards between the rich and the poor. It would aid constructive policy debate if those observers explicitly acknowledged that they are willing to see more absolute poverty in the world to assure less absolute inequality.

Another aspect of inequality that has great relevance to policy debates but received little attention in the WDR concerns *horizontal inequalities* – differences

in how ex ante similar people fare under policies and economic changes. A characterization of the distributional impacts of policy reforms that only considers the vertical redistributions involved – the differences in average gains between rich and poor, say – ignores impacts that can have both political salience and relevance to other aspects of policy. The political responses that can stall reform or create large social costs are in part horizontal in nature – between people at similar levels of living pre-reform.[6] It can be deceptive to simply average across these horizontal differences. Some reforms also combine large losses (say) for a relatively small number of people with small gains to a large number. Citizens and policymakers are likely to care about such differences. Better knowledge about such horizontal impacts can also inform the design of social protection policies that anticipate what types of households lose.

While the measures used might have been overly narrow in their focus on relative and vertical inequalities, the primary data work that was done for the WDR 2006 was nonetheless impressive (especially noting the short time frame for the work at the time). Even ten years later, this remains a valuable description of the extent of inequality in the world, including in some important non-income dimensions, such as health. The WDR's micro data set has expanded to now include harmonized variables from over 600 surveys for 120 developing countries. Unfortunately, access is currently confined to World Bank staff although I am told that there are plans for facilitating public access. This could well form the basis for a truly global compilation of micro data sets for measuring poverty and inequality.[7] Our understanding of inequality could benefit from such a global micro-data archive.

High-end inequality

Rising inequality at the top has emerged as an important policy concern, especially in the U.S. (though how much political momentum that concern will maintain remains unclear). This has been fuelled in part by the impressive data base assembled by Piketty (2014) and colleagues documenting the high and rising share held by America's top 1%. With reference to this high-end inequality, Bourguignon (2016) asks an important un-answered question: "Is too much of that inequality bad for development?"

Piketty's book is mainly about a specific source of inequality: the rising concentration of non-human wealth – Piketty's "capital" in his title, comprising physical capital, financial capital and real property. This is returning the U.S. to levels of inequality not seen since the first few decades of the 20th century. Although the evidence base is much weaker, one hears concerns about a rising concentration of non-human wealth in many developing countries today.

I think developing countries should be concerned about the damage that is possible from too much concentration of non-human wealth. The WDR

pointed to the potential for elites to undermine reforms in the collective interest. However, I would not want to see a focus on non-human wealth inequality divert attention from inequalities in human development. In most developing countries I would contend that inequality of human wealth is the more important concern for development.

Here there is a message from the 2006 WDR of continuing relevance. High returns to human capital, stemming from unequal opportunities for acquiring that capital, put upward pressure on inequality in other dimensions and represent lost opportunities for society as a whole, breeding an inequitable growth process going forward. There is a potential corrective, in that inequality of opportunity undoubtedly makes inequality of non-human wealth less acceptable in society. That helps assure the type of political consensus one often sees about reducing INOPP. There has been progress in promoting better health and education for poor people across the globe. But there is still a lot to do in assuring a reasonably level playing field.

America has been the focus of the current attention to high-end inequality. But it is also a country that today's developing countries can learn from in thinking about how best to redress inequalities in access to the means for developing human capital. From the 19th century, America was especially good at fostering high-quality public education for all, which was a key foundation for its equitable growth until the latter part of the 20th century, as Goldin and Katz (2008) have shown (though with a reversals of progress since the 1980s). Here there is a (more positive) lesson from American experience.

Notes

1. The authors suggest at one point that inequality of effort may be good for growth, though this is far from clear.
2. This interaction effect appears in the error term of the standard INOPP regressions, and is likely to be positively correlated with the regressors for circumstances; see Ravallion (2016, Chapter 3).
3. The moral difficulty in allowing a trade-off between inequality and poverty led Rawls (1971) to his famous "difference principle:" that rising inequality is only morally acceptable if it benefits the least advantaged.
4. I started voicing these concerns prior to the WDR 2006 (such as in Ravallion, 2003). This section draws in part on Ravallion (2014b).
5. Ferreira and Ravallion (2009) review the evidence on this point.
6. This is especially evident in considering trade policy, given that the heterogeneity in net trading positions in relevant markets. This point is developed further and illustrated empirically in Ravallion (2006).
7. There is already such a compilation in the form of the *Luxembourg Income Study* (LIS) although (despite some effort) this remains skewed toward rich countries (Ravallion, 2015a).

References

Bourguignon, François (2016) "Reflections on the "Equity & Development" WDR Ten Years Later," this volume.

Ferreira, Francisco, Christoph Lakner, Maria Ana Lugo, Berk Ozler (2014) "Inequality of Opportunity and Economic Growth: A Cross-Country Analysis," ECINEQ Working Paper 335.

Ferreira, Francisco and Martin Ravallion (2009) "Poverty and Inequality: The Global Context," in edited by Wiemer Salverda, Brian Nolan and Tim Smeeding (eds), *The Oxford Handbook of Economic Inequality*. Oxford: Oxford University Press.

Goldin, Claudia and Lawrence F. Katz (2008) *The Race Between Education and Technology*. Cambridge, MA: Harvard University Press.

Paes de Barros, R., F. Ferreira, F., J. Molinas Vega and J. Saavedra Chanduvi (2009) *Measuring Inequality of Opportunities in Latin America and the Caribbean*. Washington, DC: The World Bank.

Piketty, Thomas (2014) *Capital in the Twenty-First Century*. Cambridge, MA: Harvard University Press.

Ravallion, Martin (2003) "The Debate on Globalization, Poverty and Inequality: Why Measurement Matters," *International Affairs*, vol. 79, no. 4, pp. 739–54.

Ravallion, Martin (2006) "Looking Beyond Averages in the Trade and Poverty Debate," *World Development*, vol. 34, no. 8, pp. 1374–92.

Ravallion, Martin (2012) "Why Don't we See Poverty Convergence?," *American Economic Review*, vol. 102, no. 1, pp. 504–23.

Ravallion, Martin (2014a) "The Idea of Antipoverty Policy," in A.B. Atkinson and F. Bourguignon (eds), *Handbook of Income Distribution Volume 2*. Amsterdam: North-Holland.

Ravallion, Martin (2014b) "Income Inequality in the Developing World," *Science*, vol. 344, pp. 851–5.

Ravallion, Martin (2015a) "The Luxembourg Income Study," *Journal of Economic Inequality*, in press.

Ravallion, Martin (2015b) "Inequality when Effort Matters," NBER Working Paper 21394, National Bureau of Economic Research, Cambridge, Massachusetts.

Ravallion, Martin (2016) *The Economics of Poverty: History, Measurement, and Policy*. Oxford and New York: Oxford University Press.

Rawls, John (1971) *A Theory of Justice*. Cambridge, MA: Harvard University Press.

Roemer, John (1998) *Equality of Opportunity*. Cambridge, MA: Harvard University Press.

Roemer, John (2013) "Economic Development as Opportunity Equalization," *World Bank Economic Review*, vol. 28, no. 2, pp. 189–209.

World Bank (2006) *World Development Report: Equity and Development*. New York: Oxford University Press.

World Bank (2013) *World Development Indicators*. Washington, DC: World Bank.

3

Person Equivalent Headcount Measures of Poverty

*Tony Castleman, James E. Foster, and Stephen C. Smith**
George Washington University

3.1 Introduction

The most common tools for monitoring poverty are headcount measures, which evaluate a country's poverty level using the number or prevalence of poor persons in the country. Yet as emphasized by Sen (1976), headcount measures have serious limitations stemming from their inability to differentiate among the poor.[1] Large changes in incomes of the poor are ignored when the incomes stay below the poverty line, while small changes near the line can disproportionately affect measured poverty. Alternative poverty measures have been developed that address this problem by accounting for the intensity of poverty; but these measures are typically absent from policy discussions as they can be viewed as challenging for policymakers to explain intuitively, or for the public to understand.

The exclusive use of headcount measures to evaluate poverty can have significant implications for policies used to address poverty. Bourguignon and Fields (1990) demonstrate how using headcount measures encourages policy makers to ignore the poorest of the poor and focus on those with incomes just below the poverty line. Sen (1992, p. 105) contends that any government focusing solely on headcount measures "faces a strong temptation to concentrate on the *richest* among the poor, since this is the way that the number of the poor ... can be most easily reduced." A similar statement could be made for development non-governmental organizations (NGOs), international organizations, or other aid partners whose efforts are judged using poverty headcounts.[2]

*Castleman is the South Asia Representative at Catholic Relief Services and formerly Associate Director of the Institute for International Economic Policy and Associate Research Professor at George Washington University; Foster and Smith are Professors of Economics and International Affairs at George Washington University. Emails are tonyc@gwu.edu, ssmith@gwu.edu, and fosterje@gwu.edu respectively.

There is clearly a need for poverty measures that have a straightforward interpretation analogous to the headcount measures and yet appropriately reflect the intensity of poverty among the poor. This paper presents a new variety of poverty measures – called person equivalent headcount measures – to address this need. The average depth of poverty among the poor is calculated from an initial population; this benchmark "person equivalent" is used to translate between income and persons. Poverty is measured in "people space" by counting the number of person equivalents. The idea has analogies with the notion of a full time equivalent employee, which measures employment using a benchmark workweek to account for variations in the hours worked by different employees.[3]

We present two measures. The first, called the person equivalent headcount, is analogous to the headcount, or the number of poor, but instead of counting persons, it counts person equivalents. The second, called the person equivalent headcount ratio, divides through by the overall population size; it is analogous to the traditional headcount ratio, or the share of the population that is poor. Both are shown to be linked to traditional gap measures of poverty, and exhibit a number of useful properties, including two that traditional headcount measures lack: monotonicity (which requires poverty to rise if a poor person's income falls) and continuity (which requires the measure not to change abruptly with a small change in income). We note that raising the income of a minimally poor person above the poverty line will lower a person equivalent headcount by less than one; raising the income of a person from far below the poverty line to just below the poverty line will lower the person equivalent headcount by more than one. In addition, both measures are decomposable by subgroup and hence are subgroup consistent.

The new measures are related to gap measures, but differ in one key respect – their numerical values have meanings that are vivid and intuitive, as headcounts that control for the condition of the poor. Traditional headcount measures can be misleading when the conditions of the poor change dramatically. Person equivalent headcount measures benchmark the initial conditions of the poor, and then employ this standard as a measuring rod to count the number of standardized poor, or person equivalents. The picture of poverty is altered in appropriate ways: it raises the level of measured poverty when the conditions of the poor become worse; it lowers it when the average conditions are better. The extent of this alteration in practice can be captured with the elasticity of the person equivalent headcount ratio with respect to the traditional headcount ratio (or "depth elasticity").

We illustrate our methods using $1.25 a day data from PovcalNet at the World Bank, which are based on household survey data. We select two data points for each of 78 countries, one from the 1990s and a second from the 2000s, and show how the picture of poverty is altered over space and time when p. e. headcount

measures are used. Overall, the picture is one of a more rapid decline in global poverty, but with significant redistributions of its burden across regions and countries. The depth elasticities of individual countries measured are typically greater than 1.0, but with wide variation.[4] We also illustrate how poverty levels change when the benchmark population is altered, but other conclusions are not affected, including country comparisons, poverty growth rates, and depth elasticities. We conclude with a discussion of some potential topics for future research, such as applying the approach to the squared poverty gap and to the increasingly influential multidimensional poverty measures and using person equivalent headcount measures in development goals.

Section 3.2 begins with the basic definitions and notation used in the paper, while section 3.3 constructs the new measures and discusses their characteristics. The empirical examples are presented in section 3.4. Section 3.5 provides concluding remarks and future extensions.

3.2 Definitions and notation

The population size is denoted by a positive integer n, with persons represented as $i=1, \ldots, n$. The vector $x=(x_1, \ldots, x_n)$ denotes a distribution of income among the population, while a poverty line $z>0$ is used to identify when a person is poor, namely, when $x_i<z$. Let $g=(g_1, \ldots, g_n)$ be the vector of normalized gaps, where $g_i=(z-x_i)/z$ for any person i who is poor while $g_i=0$ for non-poor i. The normalized gap of a poor person expresses the shortfall $s_i=(z-x_i)$ from the poverty line as a share of the poverty line z. A poverty measure P aggregates the information in x given z to obtain an overall level $P=P(x; z)$ of poverty. A simple example is given by the poverty headcount $q=q(x; z)$ or the number of poor people in x given z, while the headcount ratio $H=q/n$ is the share of the population that is poor.

Headcount measures do not distinguish among the poor; they ignore the progress a poor person makes on the way to escaping poverty. In contrast, the poverty gap ratio $P_1=\mu(g_1, \ldots, g_n)$, which is the mean normalized gap in a population, clearly differentiates among the poor according to the depth of their poverty, and registers a decrease whenever a poor person's income rises. Notice that it can be written as $P_1=HI=HA/z$ where H is the headcount ratio, I is the income gap ratio or average normalized gap among the poor, and $A=zI$ is the average income shortfall among the poor.[5] In general, the FGT class of poverty measures can be defined for $\alpha \geq 0$ as $P_\alpha=\mu(g_1^\alpha, \ldots, g_n^\alpha)$, or the mean of the normalized gaps raised to the power α. Clearly $P_0=H$ is the headcount ratio and P_1 is the poverty gap measure, while P_2 is the FGT squared gap poverty measure that is particularly sensitive to the poorest poor and accounts for inequality among the poor.

Each of these poverty measures is relative in that it evaluates the magnitude of poverty relative to the population size, and satisfies replication invariance,

which requires a given distribution to have the same poverty level as one in which each income is replicated k times. Other measures are absolute in that they satisfy linear replication, which requires a k-replication of a distribution to have k times the poverty of the original distribution. The headcount q is one example, and the total gap $T = qA$, or the total income necessary to raise all poor persons to the poverty line, is another. Both relative and absolute measures are helpful in evaluating income poverty across populations and their subgroups.

3.3 Poverty gaps and person equivalents

Following Sen (1976), there has been a shift in the focus of poverty measurement from the identification step, by which the set of the poor are identified, to the aggregation step, by which the data are aggregated into an overall measure of poverty. Various improved aggregate measures have been proposed as replacements for the headcount q (the number of the poor) or the headcount ratio H (the share of the population that is poor). However, the simplicity of the headcount measures has continually led policymakers and applied researchers back to these crude measures.[6] Even the poverty gap measure $P_1 = HI$ which was critiqued by Sen, and popularized as part of the decomposable class of Foster, Greer, and Thorbecke (1984), is often dismissed as being too difficult for policy makers to grasp and use. Consequently, many discussions of poverty ignore significant variations in the intensity of poverty across space and time. With headcount measures, each poor person counts the same; with the poverty gap and related measures, the contribution of each poor person depends on the intensity of the poverty they experience.

In income poverty measurement, the simplest gauge of a poor person's intensity of poverty is the shortfall $s_i = z - x_i$ from the poverty line. The average depth or intensity of poverty among the poor can then be measured as A or the average shortfall among the poor. Denote the average shortfall and the poverty headcount in an initial or benchmark distribution, x^0, by A^0 and q^0 respectively. We want to measure progress when changes occur and lead to a new distribution x and its associated average intensity and headcount, A and q. For simplicity, let us initially assume that there is no population growth, so that $n = n^0$. A traditional way of assessing progress is by using the change in headcounts, with $\Delta q = q - q^0 < 0$ indicating an improvement in poverty, $\Delta q = q - q^0 > 0$ indicating a worsening, and $\Delta q = q - q^0 = 0$ suggesting that poverty is unchanged. This assessment might well be justified if the average intensity of poverty were held constant. However, if A is also changing, and the depth of poverty is seen to be of relevance in assessing progress, then headcounts by themselves can provide a misleading view of poverty. For example, if incomes of all extremely poor persons in society rose to close to, but just below the poverty line, this would be viewed using q as no progress at all. Likewise, if one person who is

marginally poor (that is, having an income just below the poverty line) became nonpoor and in the process all other poor persons become extremely poor, this would be seen as an unambiguous improvement by q. However, such a conclusion would be challenged if A were taken into account. We consider an alternative way of measuring headcounts that controls for the changes in the average intensity of poverty.

To this end, we observe that the initial intensity A^0 can be used as a measuring rod in interpreting poverty comparisons and evaluating progress. For example, suppose that the poverty line z is the traditional \$1.25 a day, and that initially there are q^0=1000 poor persons with an average shortfall of A^0 = 50¢, so that the total shortfall in the population is A^0q^0 = \$500. In the following period suppose that q is unchanged at 1000 while the average shortfall declines to A = 45¢, leading to a new total shortfall of Aq=\$450. Although the total number of poor persons is unchanged, progress has clearly been made towards reducing poverty. We can measure this improvement by dividing the new total shortfall Aq=\$450 by the original average shortfall A^0=50¢ to obtain the person equivalent (p. e.) headcount q_e=900. In words, *the person equivalent headcount q_e is the number of poor persons with a benchmark average shortfall A^0 that it would take to aggregate up to the total shortfall of Aq*. It measures the poverty gap in "people space" by using the average shortfall of poor persons as the unit of measurement.

The person equivalent headcount is the number q_e that solves A^0q_e=Aq and, hence, the person equivalent headcount is defined as

$$q_e = q\frac{A}{A^0}\,. \tag{1}$$

Progress can be gauged using the p. e. headcount, with Δq_e=q_e−q^0<0 indicating an improvement in poverty, Δq_e=q_e−q^0>0 indicating a worsening of poverty, and Δq_e=q_e−q^0=0 indicating that poverty is unchanged. The above example has Δq_e=900−1,000 = −100, which suggests that there has been progress towards reducing poverty on the order of 100 person equivalents. In contrast, the traditional headcount would indicate no progress at all. Note that the ratio A/A^0 is a transformation factor that converts the conventional headcount q into the person equivalent headcount q_e. It reinterprets the average gap using the measuring rod of the benchmark average gap. In our example, A/A^0 = 45¢/50¢ = 9/10, so that the new distribution with 1,000 poor persons and an average gap of 45¢ is viewed as having q_e=900 person equivalents.

This approach can be extended to the variable population size case to obtain a measure of poverty q_e that is absolute or independent of the number of the nonpoor. When population size varies, though, it is more traditional to report the headcount ratio, or the prevalence of the poor as a share of the overall population. Let H^0=q^0/n^0 denote the initial headcount ratio and H=q/n denote

the subsequent headcount ratio. The person equivalent headcount ratio is defined as $H_e = q_e/n$. In our previous example, suppose that the population size was initially $n^0 = 5000$ and dropped to $n = 4{,}500$. Then with $q^0 = q = 1{,}000$, we would have $H^0 = 1/5$ and $H = 2/9$ so that poverty as measured by the headcount ratio has risen, or $\Delta H > 0$. However, since $H_e = 900/4{,}500 = 1/5$, this means that the person equivalent headcount ratio is unchanged; the number of person equivalents in poverty declined at the same rate as the overall population. Note that H_e is the value that solves $A^0 n H_e = AnH$ and hence the person equivalent headcount ratio may be defined as

$$H_e = H \frac{A}{A^0} . \tag{2}$$

As before, A/A^0 is the transformation factor converting the headcount ratio H into the p. e. headcount ratio H_e. In this example, $H = 2/9$ and $A/A^0 = 9/10$, and so $H_e = 1/5$.

The above presentation has employed the average shortfall A as a measure of intensity to gauge the conditions of the poor. An alternative to the average shortfall is the income gap ratio $I = A/z$, which expresses the average shortfall as a percentage of the poverty line, rather than in monetary units, and is the intensity measure behind the poverty gap ratio $P_1 = HI$. What would change if the income gap ratio I rather than the average (monetary) gap A were used in the construction of the p. e. headcount and headcount ratio? It is easy to see that since $I/I^0 = A/A^0$, the transformation factor would remain the same, and hence the resulting person equivalent measures

$$q_e = q \frac{I}{I^0} \text{ and } H_e = H \frac{I}{I^0} \tag{3}$$

are identical to those defined in (1) and (2). Intuitively speaking, the person equivalent headcount measures compare the intensity in the later period to the benchmark intensity, and the ratio is the same whether the intensity is measured in monetary units or in poverty line units.

The poverty gap ratio $P_1 = \mu(g_1, \ldots, g_n)$ combines H and I to obtain a measure that reflects both the prevalence and intensity of poverty. It is a second indicator used to measure progress toward the poverty goal of the MDGs and is readily available on the World Bank's PovcalNet website. When published data exist on P_1, H, and n, the person equivalent measures q_e and H_e can be easily derived as follows. First find the benchmark intensity level $I^0 = P_1^0/H^0$. Then calculate the person equivalent measures as

$$q_e = nP_1/I^0 \text{ and } H_e = P_1/I^0. \tag{4}$$

In other words the person equivalent headcount ratio is found by dividing the poverty gap measure P_1 by the benchmark intensity level I^0, while the headcount further nets out the population size n.

The expression for H_e in equation (4) makes it clear that given a fixed initial average intensity level I^0, the equivalent headcount ratio H_e is proportionate to the poverty gap measure P_1. Thus H_e is a poverty measure that evaluates distributions in the same way as P_1, but has an alternative interpretation as the number of person equivalents per capita. It shares the same properties as P_1, including symmetry, population replication, the focus axiom, monotonicity, continuity and decomposability.[7] In contrast, the original headcount ratio H does *not* satisfy monotonicity and continuity; it ignores all improvements in the conditions of the poor that do not result in a crossing of the poverty line, but registers a discrete change when a poor person does cross. For instance, if a program targeting the ultra-poor successfully lowered the income shortfalls of the ultra-poor by 90 percent, it would be regarded by H with indifference – the prevalence of poverty has simply not changed. The program's progress would be revealed if the p. e. headcount ratio H_e were used. In an analogous fashion, equation (1) reveals that the person equivalent headcount q_e is proportionate to the total income gap $T=qA$ and hence evaluates distributions in a similar way. The properties satisfied by q_e (and T) include symmetry, linear replication, the focus axiom, monotonicity, continuity and additivity, all of which are satisfied by the headcount measure q apart from monotonicity and continuity.[8] The monotonicity axiom ensures that q_e reflects the changes in the intensity of poverty even when q is unchanged.

We have described two new measures of poverty – the person equivalent headcount q_e and the person equivalent headcount ratio H_e – that evaluate poverty in "people space" with the help of a transformation factor based on the average depth or intensity of poverty in a benchmark period. If average depth falls below benchmark, the person equivalent headcount measure will be lower than its respective traditional headcount measure; if average depth rises above benchmark, the person equivalent headcount measure will be higher than the traditional headcount measure. To recap, when a person who was poor in an initial period crosses the poverty line, the impact on a person equivalent headcount measure depends on the depth of the person's poverty in the prior period: If the initial income was slightly below the poverty line, it would have a small effect on a person equivalent headcount measure, while if the income was well below the poverty line, it would have a larger effect.

In certain contexts, we might be mainly interested in evaluating the percentage change in poverty through time. To evaluate the inclusiveness (or pro-poorness) of growth, for example, the rate of reduction in poverty can be divided by the growth rate in per capita income to derive the growth elasticity of poverty – a measure of how well the economy is converting income growth into poverty alleviation. The traditional growth elasticity uses the headcount ratio H; the

person equivalent headcount ratio H_e could also be used. A related statistic compares the growth elasticities of H_e and H (or equivalently, the growth rates of H_e and H) to obtain the depth elasticity of poverty $\varepsilon = \%\Delta H_e/\%\Delta H$, which indicates how progress in the headcount ratio is translating into progress in the person equivalent headcount ratio.[9]

The above discussion applies to evaluations of progress over time for a given population. The analysis can be readily extended to comparisons between subgroups defined according to geographical location, demographic characteristic, or some other parameter. For example, person equivalent headcount measures for different countries can be constructed and compared using the global $1.25 a day poverty standard using a benchmark intensity level drawn from world data. Alternatively, interest might be regional in scope, in which case p. e. headcount comparisons could be made across countries within the region using a regional benchmark. Both examples divide an overall distribution x into subgroup distributions and evaluate subgroup poverty using a person equivalent headcount measure benchmarked for the overall population. For simplicity of notation, let us focus on the two-subgroup case where the distribution can be written as $x=(a, b)$ for subgroup distributions a and b; the same logic would apply to the many subgroup case. The benchmark intensity A^0 is obtained from an initial distribution x^0 drawn from the same general population as x, but potentially at an earlier time period. The case $x^0=x$ corresponds to an analysis of poverty over space (i.e., across subgroups); the case where x^0 is drawn from an earlier time period leads to comparisons over space and time.

Let q^a and A^a be the headcount and average intensity levels for distribution a, and let q^b and A^b be the associated values for distribution b. By the additivity of the headcount measure q we know that

$$q=q^a+q^b \tag{5}$$

or the overall headcount in x is the sum of the respective headcounts in a and b. Given the benchmark level of intensity A^0, define the person equivalent headcount for distributions a and b by $q_e^a = q^aA^a/A^0$ and $q_e^b = q^bA^b/A^0$. Since $qA=q^aA^a+q^bA^b$, it immediately follows from dividing through by A^0 that

$$q_e = q_e^a+q_e^b \tag{6}$$

so that the overall person equivalent headcount is the sum of the respective person equivalent headcounts in a and b. As compared to the crude headcount q, the p. e. headcount $q_e=qA/A^0$ is higher or lower depending on whether the intensity in x is, respectively, higher or lower than the benchmark level. Equation (6) then provides the breakdown of the person equivalent headcount across the subgroups.

If $x = x^0$, the overall person equivalent headcount in x reduces to $q_e = q\left(\frac{A^0}{A^0}\right) = q$, the traditional headcount, so that q_e from equation (6) is equal to q from equation (5). The decomposition in (6) is an alternative poverty breakdown to (5) that accounts for the differential depth of poverty experienced by people in the two regions. Using person equivalent headcounts can be interpreted by imagining that there is a redistribution of the population of poor people across the subgroups, with the higher intensity subgroup gaining person equivalents and the lower intensity subgroup losing. If there are two periods, then equation (6) can also be applied to the second period using the base intensity level A^0 and comparing across time to obtain the changes in each subgroup and overall. It is easy to show that $\Delta q_e = \Delta q_e^a + q_e^b$, so that the change in the overall p. e. headcount is the sum of the respective changes in p. e. headcounts for the two subgroups.

From equation (5), or by the decomposability property of the headcount ratio, it follows that

$$H = (n^a/n)H^a + (n^b/n)H^b \tag{7}$$

and so the overall headcount ratio is a population-share weighted average of the subgroup headcount ratios. An analogous argument using equation (6) or decomposability for P_1 yields

$$H_e = (n^a/n)H_e^a + (n^b/n)H_e^b \tag{8}$$

which is the decomposition formula for the person equivalent headcount ratio. If applied to the original distribution, H_e becomes H so that (8) provides an alternative breakdown of the headcount ratio accounting for the intensities of poverty in the two groups. Equation (8) can also be used over time to link progress in the person equivalent headcount ratio to progress at the subgroup level.

The interpretations of q_e and H_e depend centrally on the benchmark level of intensity, and hence the time and region from which it is drawn. The benchmark is an average value in a region (which could be a particular country, a collection of countries, or the world) at a given point in time. A poor person with smaller than average income gap will account for less than one person equivalent; a poor person with a larger than average gap will add more than one. If a subgroup contains many poor people who are deeply poor, and the regional benchmark is sufficiently small, the number of person equivalents in the subgroup could well exceed the subgroup population, leading to a p. e. headcount ratio beyond the usual bounds.[10] Likewise, a region with a large number of poor persons just below the poverty line could record a much lower p. e. headcount

than its traditional headcount, particularly if the regional benchmark is high. In any case, the subgroup levels stay in proportion with one another even as the benchmark changes.

The technology of person equivalent headcounts is well suited for formulating development goals and targets that go beyond crude headcount measures and have a natural starting time from which to benchmark. The decomposition formula would support multilevel analyses at different granularities, from the global level down to an individual household. The methods could be adapted to different purposes by altering the region used in benchmarking. For example, to monitor global progress, the average intensity across the globe in the starting period could be used as the benchmark. A regional development bank might be interested only in reporting progress in its target area and could use a benchmark from a geographic region such as Sub-Saharan Africa. Moreover, a country could monitor its own progress using the initial countrywide average intensity as a benchmark. And, as noted above, even when the benchmark is varied to reflect the different purposes and scope of analyses, the results are quite consistent. The person equivalent headcount measures obtained using one benchmark are proportional to those obtained with another, preserving their relative magnitudes across space, and their growth rates across time. There is no possibility of misaligned incentives for the parties conducting the analyses at different levels.

For example, suppose that the initial average intensity is $I_C^0 = 0.50$ in a country, $I_R^0 = 0.75$ in the region, and $I_W^0 = 0.25$ in the world. If the country's poverty gap ratio were 0.12 in the initial period and 0.06 in a subsequent period, the person equivalent headcount ratios as reckoned using the country benchmark would initially be 0.24 and subsequently 0.12. If instead the regional benchmark were used, then the person equivalent headcount ratios would be 0.16 and 0.08, respectively, or $I_C^0/I_R^0 = 2/3$ times these values, while at the world benchmark, the values would be 0.48 and 0.24, respectively, or $I_C^0/I_W^0 = 2$ times as large as the values using the country benchmark. Note that the trends in poverty for a given country are consistent irrespective of which benchmark is employed. Moreover, if the goal were to lower the person equivalent headcount ratio by 50 percent of the initial value, all three levels would track progress consistently and would meet the goal at the same time. In this sense, the goal would be robust to the choice of benchmark.[11]

3.4 An illustration: global poverty

The person equivalent approach to evaluating poverty is illustrated using poverty data published by the World Bank on PovcalNet, which is based in turn on underlying household survey data.[12] In order to gauge progress over time, we restrict consideration to countries for which data exist for at least one year

in an initial range of 1992–2000 and one year in the later range of 2005–10. A total of 78 countries from six regions satisfy this criterion. We construct a "developing world" made up of these countries, and extract the $1.25 a day figures to create country poverty data for two time periods. Decomposition formulas allow headcount ratios and the poverty gap ratios to be calculated for the developing world and for regions. We apply the formula $A^0 = zP_1^0/H^0$ (or equivalently $I^0=P_1^0/H^0$) to data from the initial period to derive the appropriate benchmark level, which in turn is used to produce person equivalent headcounts and person equivalent headcount ratios.

Table 3.1 reports the poverty statistics for our full 78-country sample over the two periods using the global benchmark level of A^0=39.5¢ per day (or equivalently I^0=0.316). The global headcount q dropped by 512 million persons during this period, and since the average income shortfall among the poor also declined by more than 4¢, to A=35.2¢ per day, the drop in person equivalents was 625 million, 113 million more than the drop in headcount. The conventional headcount ratio fell 44 percent from H^0=0.36 to H=0.20 between the two periods, while the person equivalent headcount ratio decreased by 50 percent from H_e^0=0.36 to H_e^1=0.18, once again reflecting the decrease in average depth of poverty. The global depth elasticity was about 1.1, indicating that for every one percent drop in the headcount ratio there was a 1.1 percent decline in the person equivalent headcount ratio.

Table 3.2 explores regional poverty levels and trends using the same global benchmark.[13] The move to person equivalent headcounts from traditional headcounts results in increases in poverty rates in Sub-Saharan Africa (SSA) and Latin America and the Caribbean (LAC) and decreases in the other regions; this is true for both periods. A comparison of South Asia (SA) and SSA is particularly informative. The poverty headcount in SA was initially 608 million, far higher than in SSA at 260 million, and SA continued to dominate SSA in the numbers of poor people by a wide margin of over 180 million in the second period, even as the headcount fell in SA and rose in SSA. However, when viewed through the lens of person equivalent headcounts, the initial levels of the two regions are seen to be much closer and, in the second period, SSA actually overtakes SA by more than 35 million person equivalents. Incorporating the depth of poverty paints a rather different picture of poverty and progress in the two regions than headcount alone.

Headcount ratios take into account the differential population sizes across countries and through time. The data for H show strong declines for all regions, with the success story of East Asia and the Pacific (EAP) being represented by a sharply falling value of H (from 0.38 to 0.11). Notice that the decline in the headcount ratio in LAC (from 0.09 to 0.05) is also impressive in percentage terms. Indeed, the data on person equivalent headcount ratios also show progress in H_e for both regions; but differences in average intensity shift the values for EAP

Table 3.1 Person equivalent headcount measures: full sample (1990s global benchmark)

	Range of years	Population n (millions)	Headcount q (millions)	Headcount ratio H	Person-equivalent headcount q_e (mill.)	Person-equivalent headcount ratio H_e	% change in H	% change in H_e	Depth elasticity
World (78 countries)	1992–2000	4,231	1,547	.36	1,547	.36	–44.3%	–50.4%	1.1
	2005–2010	5,189	1,035	.20	922	.18			

Note: Benchmark is global average income shortfall in 1992–2000 period: 39.5¢ per day.

Table 3.2 Person equivalent headcount measures by region (1990s global benchmark)

Region	Range of years	Population n (millions)	Headcount q (millions)	Headcount ratio H	Person-equivalent headcount q_e (mill.)	Person-equivalent headcount ratio H_e	% change in H	% change in H_e	Depth elasticity
East Asia & Pacific	1992–97	1,635	615	.38	575	.35	–71%	–79%	1.1
	2007–10	1,842	201	.11	134	.07			
Europe & Central Asia	1993–2000	399	15	.04	12	.03	–74%	–83%	1.1
	2007–10	402	4	.01	2	.01			
Latin America & Caribbean	1993–99	457	43	.09	60	.13	–49%	–48%	0.97
	2005–10	535	26	.05	38	.07			
Middle East & North Africa	1994–98	179	7	.04	4	.02	–35%	–25%	0.73
	2005–10	213	5	.02	3	.02			
South Asia	1993–96	1,210	608	.50	537	.44	–37%	–49%	1.3
	2007–10	1,553	489	.32	354	.23			
Sub-Saharan Africa	1992–2000	441	260	.59	359	.81	–16%	–23%	1.4
	2005–10	625	308	.49	390	.62			

Note: Benchmark is global average income shortfall in 1992–2000 period: 39.5¢ per day.

down and LAC up, relative to traditional headcount ratios, with the result that both regions reach the same level of H_e (namely, 0.07) in the second period. Returning to the case of SSA and SA, the initial values for H are not dissimilar (at 0.59 and 0.50, respectively). However, regional differences in intensity generate a wide divergence in the initial period's values of H_e for SSA and SA (0.81 and 0.44, respectively), while differences in progress accentuate this further so that the final period H_e values in SSA and SA are, respectively, 0.62 and 0.23. Depth elasticities were greater than 1.0 in four regions (SSA, SA, ECA, and EAP), indicating faster reduction in H_e than H. LAC had an elasticity of 0.97, indicating similar rates of reduction in the two headcount measures. The Middle East and North Africa (MENA) had a depth elasticity of 0.73. Apparently, the reduction in H was accompanied by an increase in the depth of poverty, leading to a smaller improvement in the p. e. headcount ratio H_e than in H.

Table 3.3 contains poverty statistics for 11 of the 78 countries using the same global benchmark as above. We examine how the levels of H and H_e in a country differ from one another and how this alters the rate at which poverty changes. Four of the countries (China, India, Egypt, and South Africa) follow the lead of the global figures and the EAP and SA regions, by having smaller values for H_e than H in both periods; four other countries (Bolivia, Brazil, Kenya, and Mozambique) follow the LAC and SSA regions by having higher H_e values in both periods, indicating that the average depth is greater than the benchmark levels for both periods. In the three remaining countries (Vietnam, Nepal, and Niger), the level of H_e is higher than H in the initial period and lower than H in the second, indicating faster poverty reduction in these countries when depth of poverty is taken into account than when only the prevalence of poverty is measured. Indeed, all countries except for Brazil and Egypt exhibit a higher rate of change in H_e than in H, and thus have depth elasticities that exceed 1.0. For example, Nepal's conventional headcount ratio declined 4.4 percent per year and its p. e. headcount ratio declined by 5.4 percent per year, leading to an elasticity of 1.2. Two countries, Kenya and Bolivia, saw poverty rise between the two periods; large increases in their headcount ratios were magnified even further by their elasticities of 1.5 and 2.3, respectively. A depth elasticity greater than one means that the average depth of poverty worsened in countries where the prevalence of poverty worsened and improved in countries where the prevalence of poverty improved. For Brazil and Egypt the picture is different. Brazil's conventional headcount ratio declined more quickly (3.7 percent per year) than its p. e. headcount ratio (3.0 percent per year) resulting in a depth elasticity of 0.8; Egypt's p. e. headcount ratio increased slightly (0.7 percent per year) while its conventional headcount ratio declined (2.5 percent per year), yielding a case where the depth elasticity takes on a negative value of –0.3.

Table 3.4 presents the second period data for countries in Sub-Saharan Africa using a contemporaneous regional benchmark of A^0 = 50.0¢ per day (the

Table 3.3 Person equivalent headcount measures for selected countries (1990s global benchmark)

Country	Year	Population n (millions)	Headcount q (millions)	Headcount ratio H	Person-equivalent headcount q_e (mill.)	Person-equivalent headcount ratio H_e	Annual % change in H	Annual % change in H_e	Depth Elasticity
Bolivia	1993	7	0.6	.09	0.8	.11	1.5%	3.5%	2.3
	2008	10	1.0	.16	1.6	.17			
Brazil	1995	162	16	.11	21	.13	–3.7%	–3.0%	0.8
	2009	193	9	.05	15	.08			
China	1996	1,218	455	.37	427	.35	–5.4%	–5.8%	1.1
	2010	1,338	123	.09	86	.06			
Vietnam	1993	68	44	.64	51	.74	–4.8%	–5.5%	1.1
	2008	85	14	.17	10	.12			
India	1993	921	455	.49	395	.43	–2.1%	–2.8%	1.3
	2009	1,190	388	.33	282	.24			
Nepal	1995	21	14	.68	17	.81	–4.4%	–5.4%	1.2
	2010	27	6	.24	4	.16			
Egypt	1995	61	1.5	.025	.66	.011	–2.5%	0.7%	–0.3
	2008	76	1.3	.017	.88	.012			
Kenya	1994	27	8	.29	8	.30	4.6%	7.1%	1.5
	2005	36	16	.43	19	.54			
Mozambique	1996	16	13	.81	21	1.3	–2.2%	–3.3%	1.5
	2007	23	14	.61	19	.82			
Niger	1994	9	7	.78	11	1.2	–3.5%	–5.3%	1.5
	2007	14	6	.42	5	.37			
South Africa	1995	39	8	.21	6	.17	–2.6%	–4.1%	1.6
	2008	50	7	.14	4	.07			

Note: Benchmark is global average income shortfall in 1992–2000 period: 39.5¢ per day.

Table 3.4 Countries in Sub-Saharan Africa (2000s regional benchmark)

Country	Year	Population n (millions)	Headcount q (millions)	Headcount ratio H	Person-equivalent headcount q_e (mill.)	Person-equivalent headcount ratio H_e
Burkina Faso	2009	15	6.7	.44	5.5	.37
Burundi	2006	8	6.5	.81	7.3	.91
Cameroon	2007	19	5.2	.28	3.5	.18
Central African Republic	2008	4	2.6	.63	3.3	.78
Cote d'Ivoire	2008	18	6.4	.35	5.8	.32
Ethiopia	2010	87	34	.39	23	.26
Ghana	2005	21	6.1	.29	5.3	.25
Guinea	2007	10	4.0	.39	3.3	.33
Kenya	2005	36	16	.43	15	.42
Madagascar	2010	21	18	.88	26	1.2
Malawi	2010	15	10.8	.72	12.9	.86
Mali	2010	14	7.1	.51	5.8	.41
Mauritania	2008	3.4	.80	.23	.58	.17
Mozambique	2007	23	14	.61	15	.65
Niger	2007	14	6.0	.42	4.2	.29
Nigeria	2009	155	96	.62	107	.67
Rwanda	2010	11	6.8	.63	7.2	.66
Senegal	2005	11	3.8	.34	3.0	.27
South Africa	2008	50	6.8	.14	2.8	.06
Swaziland	2009	1.2	.46	.40	.45	.38
Tanzania	2007	41	28	.68	29	.70
Uganda	2009	33	12	.38	10	.30
Zambia	2010	13	9.8	.74	14	1.0

Note: Benchmark is Sub-Saharan Africa region's average income shortfall in 2005–2010: 50.0¢ per day.

average income shortfall in the 2005–2010 period in the SSA countries). Note the wide range in the person equivalent headcount ratios (from 0.06 to 1.2) as compared to the conventional headcount ratios (ranging from 0.14 to 0.88), suggesting that the intensity is higher than the regional average for some countries with high prevalence of poverty, and lower than average for some countries with low prevalence. Cameroon, Niger, and South Africa, for example have much lower p. e. headcounts than traditional headcounts, while Madagascar and Zambia have much higher p. e. headcounts. Note that the relative picture across SSA countries would be the same if a different benchmark were used, such as the regional benchmark computed from period one data or the global benchmark used above: q_e and H_e would simply be shifted proportionally to reflect the new standard.

Table 3.5 shows how an individual country, Niger, can use its own base year average income shortfall (61.7¢ per day in 1994) to benchmark its progress in combating poverty. Between 1994 and 2007, the number of individuals in Niger earning less than $1.25 per day dropped from 6.9 million to 6.0 million, a modest decline. The person equivalent headcount dropped from 6.9 million to 3.4 million during the same period, a very large decrease. The person equivalent headcount ratio decreased at an annualized rate of -5.3 percent per year and, as discussed above, this percentage change would be the same regardless of whether the benchmark used is the country's average income shortfall, the region's, or the world's. However, for its own internal assessment, a country may focus on the locally benchmarked figures, while knowing the findings will be consistent with globally, regionally, or even arbitrarily benchmarked figures.

3.5 Conclusions

In this paper we have presented person equivalent headcount measures, which like traditional headcount measures are evaluated in "people space"; but instead of counting poor persons, these measures count person equivalents, as benchmarked by the average depth of poverty in a given place and time. The resulting measures are intuitive to explain and easy to calculate, but at the same time satisfy monotonicity and are sensitive to the depth of poverty like the poverty gap measures to which they are related. By explicitly accounting for the conditions of the poor, they remove the incentive to focus on the least deprived segments of poor populations. When an extremely poor person escapes poverty, this has a greater impact on the measures than when a marginally poor person crosses the poverty line. And if a poor person makes good progress towards escaping poverty, but has not yet crossed the poverty line, this is regarded by person equivalent headcount measures as a positive achievement rather than something to be ignored.

We then applied our new measures to $1.25-a-day global poverty data to show how they paint a different picture of poverty and progress than conventional

Table 3.5 Person equivalent headcount measures in Niger (1994 country benchmark)

Year	Population n (millions)	Headcount q (millions)	Headcount ratio H	Person-equivalent headcount q_e (mill.)	Person-equivalent headcount ratio H_e	Annual % change in H	Annual % change in H_e	Depth elasticity
1994	8.9	6.9	.78	6.9	.78	–3.5%	–5.3%	1.5
2007	14.2	6.0	.42	3.4	.24			

Note: Benchmark is Niger's average income shortfall in 1994: 61.7¢ per day.

headcount measures. For example, person equivalent headcounts are much higher in Sub-Saharan Africa and in Latin America and the Caribbean than traditional headcounts, and lower in South Asia and East Asia and the Pacific. Interestingly, SSA joins the SA and EAP regions in registering faster percentage declines in person equivalent headcount ratios than in traditional headcount ratios, while in LAC the rates of decline in the two measures are quite similar. An analysis by country likewise provides new insights into country experiences, with some larger countries like China, India, and Brazil closely matching their regional results and others like South Africa departing widely from their regional picture.

Most of the countries we examined have a depth elasticity – or the percentage change in p. e. headcount ratio over the percentage change in the traditional headcount ratio – that is greater than one. South Africa lowered its headcount ratio a great deal, but with a depth elasticity of $\varepsilon = 1.6$, the performance in terms of person equivalents was even more impressive. Kenya had a similar depth elasticity of $\varepsilon = 1.5$, but since the headcount ratio rose, the elasticity indicates an even more dramatic increase in p. e. headcount ratio. Other countries like Brazil with $\varepsilon < 1.0$ had their improvements in p. e. headcount measures muted as compared to headcount ratios. Two countries – Egypt and Mauritania – exhibited negative depth elasticities, but they also had very small changes in headcount (and p. e. headcount) ratios. We illustrated how the benchmark underlying person equivalent measures is altered by using different geographic areas, from a global to regional or even country level. Different benchmarks result in different values for a country's person equivalent headcount measures, but since countries are affected proportionally, rankings are consistent and growth rates are unchanged. This consistency makes the person equivalent headcount ratio especially appropriate for use in multilevel development goals.[14]

The person equivalent headcount measures provide an intuitive way of incorporating information on the depth of poverty, but some might contend that they move too far afield from traditional headcount measures. Indeed, if the condition of a person changed discontinuously as the poverty line is crossed, it could make sense to retain this feature in a measure of poverty. One approach could be to construct "hybrid" measures such as $q_\lambda = \lambda q + (1-\lambda)q_e$ or $H_\lambda = \lambda H + (1-\lambda)H_e$ where $\lambda \in (0,1)$ represents the extent one believes that the discontinuity (and hence q or H) is important.[15] Now poverty is evaluated not only by counting the poor or counting person equivalents, but through a compromise between the two perspectives. This practical approach, however, can reintroduce an incentive to focus on the minimally poor – at least in the simplified perfect information scenario of Bourguignon and Fields (1990). It would be interesting to see whether the hybrid measures can provide benefits in other more realistic environments, such as when information asymmetries (say, between policymakers and aid workers) play a significant role.

One possible critique of the presentation up to now is its exclusive focus on monetary poverty. As emphasized in the 2000 *World Development Report* of the World Bank, poverty goes beyond monetary resources: it depends centrally on other key dimensions that should also be included when identifying the poor and measuring poverty. Of course all of the above measurement technology will apply directly to any other (cardinal) single dimensional variable (e.g. schooling or nutrition), thus identifying persons who are deprived in that variable and measuring their levels of deprivation.[16]

However, it is now generally recognized that the multiple dimensions must be simultaneously observed in order to identify who is poor and to evaluate how poor they are. And this typically requires expanded data and a new measurement technology. Distributions are now matrices, the single poverty line becomes a vector of "deprivation cutoffs," and an overall measure is constructed by aggregating across dimensions for the persons identified as poor. Unlike the unidimensional case, identification is not a simple matter in the multidimensional context; indeed, most theoretical presentations do not provide a practical method for identifying the multidimensionally poor, but instead fall back on a "union" approach, equating poverty with being deprived in any dimension. Moreover, the indicators available for multidimensional poverty analysis are often ordinal, rendering many theoretical solutions to the identification and aggregation steps inapplicable. The multidimensional headcount ratio H_m is one index that works well with ordinal data; however, both it and the multidimensional headcount $q_m=nH_m$ suffer from the flaws of their unidimensional cousins as highlighted in this paper.

The challenge, then, of multidimensional poverty measurement has been to solve the identification and aggregation steps in a way that is consistent with ordinal data, but goes beyond crude headcount measures. One methodology that does this is found in Alkire and Foster (2011).[17] A person is deprived in a given dimension if the achievement level is below a deprivation cutoff for the dimension. Each deprivation has a "value" and a person is poor or not depending on the extent of the person's multiplicity of deprivations, as measured by the deprivation count or sum of these values (where the maximum sum of all values is fixed at d, the number of dimensions). For example, if each deprivation has the same value, then the deprivation count is the number of deprivations the person is experiencing at the same time. A person is poor if the deprivation count meets or exceeds a poverty cutoff set between 0 and d. A poor person's intensity of poverty is measured as the deprivation count divided by its maximum d. The average intensity, denoted A_m, is the sum of the intensities of the poor divided by their number q. The adjusted headcount ratio is then given by $M_0=H_m A_m$.

Note that the form of this measure is entirely analogous to that of the poverty gap ratio $P_1=HI$, which underlies the person equivalent headcount measures for monetary poverty. Could our technology be applied in the multidimensional

case to transform the adjusted headcount ratio into a multidimensional person equivalent headcount ratio? If so, then it could offer helpful interpretations for the many applications of M_0 in common use, including official measures in several countries and the Multidimensional Poverty Index (MPI) published in the annual Human Development Report by the United Nations.[18] This would be a useful direction to pursue in future work.

In addition, we could also consider monetary poverty measures that stress the conditions of the poorest poor and take into account inequality among the poor. The distribution sensitive measures of Sen (1976), Watts (1968) and Foster, Greer and Thorbecke (1984) all have this characteristic and, as noted by Bourguignon and Fields (1990), they provide a positive incentive for focusing on the poorest poor first. Would it be possible to construct person equivalent headcount measures for each that would appropriately reflect inequality?

Consider the case of the squared poverty gap P_2 of Foster, Greer and Thorbecke. As noted above, P_2 places greater weight on persons who are further below the poverty line by squaring the normalized gaps g_i before averaging. Clearly, $P_2=HI_2$ where H is the headcount ratio and I_2 is the average of g_i^2 among the poor (an alternative intensity measure that accounts for inequality among the poor). Letting I_2^0 be its benchmark level, we can define $H_e'=P_2/I_2^0$ and $q_e'=nH_e'$ as the person equivalent measures associated with P_2. An associated "severity elasticity" could evaluate the elasticity of H_e' with respect to H, or $\%H_e'/\%H$.

Preliminary results using PovcalNet data suggest how accounting for distribution sensitivity by using P_2 influences the picture of global poverty. Global totals are virtually unchanged (see Table 3.6). However, the regional picture becomes even more pronounced, with H_e' and q_e' moving further in the directions taken by H_e and q_e (see Table 3.7). Now the EAP region begins with 524M person equivalents and ends up with only 94 million. In contrast, SSA rose from 456 million in the 1990s to 471 million in the 2000s – almost twice the number of person equivalents in SA and five times the number in EAP. The severity p. e. headcount ratio (H_e')[19] for LAC of 0.11 is more than double that of EAP (0.05). Applying the person equivalent approach to distribution sensitive measures of poverty is an interesting topic for future work.

We have emphasized the suitability of the person equivalent technology for defining and tracking multilevel development goals. We now conclude with a brief discussion of its relevance to the Millennium Development Goals (MDGs) and the post- 2015 agenda. Recall that the main indicator for the poverty portion of Goal 1 of the MDGs has been the $1.25 a day headcount ratio H. The poverty gap ratio P_1 is also listed as a complementary indicator, but for reasons of simplicity has largely been absent from all but the most technical discussions. For example, in its assessment of progress in the MDGs, the United Nations (2014) presents only headcounts or headcount ratios as indicators of success or failure in reaching poverty goals. The World Bank's (2010) assessment uses P_1,

Table 3.6 Global person-equivalent poverty including squared gaps (1990s global benchmark)

	Years	n (mill.)	q (mill.)	H	q_e (mill.)	H_e	q_{e2} (mill.)	H_{e2}	%ΔH	%ΔH_e	%ΔH_{e2}	Depth elasticity	Severity elasticity
World (78 countries)	1992–2000	4,321	1,547	.36	1,547	.36	1,547	.36	–44.3%	–50.4%	–51.8%	1.1	1.2
	2005–2010	5,189	1,035	.20	922	.18	897	.17					

Notes: Benchmark for q_e and H_e is global average income shortfall in 1992–2000 period: 39.5¢ per day
Benchmark for q_{e2} and H_{e2} is global average squared income shortfall in 1992–2000: 22¢^{2} per day, equivalent to a gap of 47¢ per day..

Table 3.7 Person equivalent poverty among regions including squared gaps (1990s global benchmark)

Region	Years	n (mill.)	q (mill.)	H	q_e (mill.)	H_e	q_{e2} (mill.)	H_{e2}	%ΔH	%ΔH_e	%ΔH_{e2}	Depth elasticity	Severity elasticity
East Asia & Pacific	1992–97	1,635	615	.38	575	.35	524	.32	–71%	–79%	–84%	1.1	1.2
	2007–10	1,842	201	.11	134	.07	94	.05					
Europe & Central Asia	1993–2000	399	15	.04	12	.03	11	.03	–74%	–83%	–80%	1.1	1.1
	2007–10	402	4	.01	2	.01	2.3	.01					
Latin America & Caribbean	1993–99	457	43	.09	60	.13	90	.20	–49%	–48%	–46%	0.97	0.93
	2005–10	535	26	.05	38	.07	59	.11					
Middle East & North Africa	1994–98	179	7	.04	4	.022	2.8	.016	–35%	–.25%	–5%	0.73	.14
	2005–10	213	5	.02	3	.016	3.2	.015					
South Asia	1993–96	1,210	608	.50	537	.44	464	.38	–37%	–49%	–55%	1.3	1.5
	2007–10	1,553	489	.32	354	.23	268	.17					
Sub-Saharan Africa	1992–2000	441	260	.59	359	.81	456	1.0	–16%	–23%	–27%	1.4	1.7
	2005–10	625	308	.49	390	.62	471	.75					

Notes: Benchmark for q_e and H_e is global average income shortfall in 1992–2000 period: 39.5¢ per day.
Benchmark for q_{e2} and H_{e2} is global average squared income shortfall in 1992–2000: $22¢^2$ per day, equivalent to a gap of 47¢ per day.

but only to help explain why poorer countries might have low growth elasticities and slower progress in reducing *H*. We would argue that depth should be included into the mix, both when evaluating the initial distribution of poverty and in monitoring the progress of countries. The person equivalent technology provides a simple and intuitive way of doing just this.

Ending extreme monetary poverty – interpreted by the UN and World Bank as reducing the $1.25 a day headcount ratio to no more than 3 percent of the population – has emerged from the post-2015 discussion as a possible poverty goal. Assuming that the global population will be nine billion, this could translate to a headcount of 270 million left behind. If this goal were achieved and only 270 million people remained below the $1.25 per day poverty line, it stands to reason that this group could contain some of the most deeply deprived, difficult-to-reach persons on earth. There is nothing in the goal that would prevent them from having an average depth of poverty that is twice the average depth of poverty among the poor in 2015. Using the 2015 average depth as the benchmark, this would mean a person equivalent headcount of over a half a billion. Should this really be seen as an end to extreme poverty? Restating the goal in terms of person equivalent headcount ratios removes the ambiguity about the conditions of those left behind. Monitoring progress with these measures ensures that the depth of poverty is also being evaluated through time. Assessing initial conditions using person equivalent headcount measures presents a more complete guide to the challenges that lie ahead.

Appendix: Countries and Years

Cambodia	1994	2009	Latvia	1996	2009
China	1996	2010	Lithuania	1998	2008
Indonesia	1996	2010	Macedonia	2000	2008
Lao PDR	1997	2007	Moldova Rep	1997	2010
Philippines	1994	2009	Poland	1996	2010
Thailand	1994	2010	Romania	1998	2010
Vietnam	1992	2008	Russian Fed	1996	2006
Albania	1996	2008	Tajikistan	1999	2009
Armenia	1998	2010	Turkey	1994	2010
Azerbaijan	1995	2008	Ukraine	1995	2010
Belarus	2000	2010	Argentina*	1995	2010
Croatia	1998	2008	Bolivia	1993	2008
Georgia	1996	2010	Brazil	1995	2009
Hungary	1998	2007	Chile	1994	2009
Kazakhstan	1996	2010	Colombia	1996	2010
Kyrgyz Rep.	1993	2010	Costa Rica	1995	2009

(*continued*)

Appendix: Countries and Years Continued

Dominican R	1996	2010	Burkina Faso	1994	2009
Ecuador	1999	2010	Burundi	1992	2006
El Salvador	1995	2009	Cameroon	1996	2007
Guatemala	1998	2006	Central Afr R	1992	2008
Honduras	1996	2008	Côte d'Ivoire	1995	2008
Mexico	1994	2010	Ethiopia	1995	2010
Nicaragua	1993	2005	Ghana	1998	2005
Panama	1995	2010	Guinea	1994	2007
Paraguay	1995	2010	Kenya	1994	2005
Peru	1997	2010	Madagascar	1993	2010
Uruguay*	1995	2005	Malawi	1997	2010
Venezuela	1995	2006	Mali	1994	2010
Egypt Arab R	1995	2008	Mauritania	1995	2008
Iran Islamic R	1994	2005	Mozambique	1996	2007
Jordan	1997	2010	Niger	1994	2007
Morocco	1998	2007	Nigeria	1996	2009
Tunisia	1995	2010	Rwanda	2000	2010
Yemen Rep	1998	2005	Senegal	1994	2005
Bangladesh	1995	2010	South Africa	1995	2008
India	1993	2009	Swaziland	1994	2009
Nepal	1995	2010	Tanzania	1991	2007
Pakistan	1996	2007	Uganda	1996	2009
Sri Lanka	1995	2009	Zambia	1996	2010

*Argentina and Uruguay data are urban only.

Notes

1. Sen's (1976, p. 219) critique begins with the observation that headcount measures ignore the poverty depth: "An unchanged number of people below the 'poverty line' may go with a sharp rise in the extent of the short-fall of income from the poverty line." He also attacks headcount measures for ignoring the distribution of income among the poor.
2. In a 2012 presentation, Steve Radelet, former Chief Economist with USAID, urged the development community to look beyond headcount measures, which ignore progress that takes place below the poverty line (Risley, 2012). Our paper was written in response to his message and the real world examples he described in subsequent conversations.
3. See also the related notions of *adult equivalent incomes* commonly used in distribution analysis, the *equally distributed equivalent income* of Atkinson (1970), or *adult equivalent labor* as in Basu and Pham's (1998) model of child labor.
4. A caveat is that some of the countries that we could not include due to lack of data have been estimated otherwise to have high average poverty depth; these include Afghanistan, Congo, Guinea-Bissau, Eritrea, Haiti, Liberia, Sierra Leone, and Zimbabwe. Several of the omitted countries have a history of violent conflict – perhaps a reason for missing data, but also a likely cause of poverty.
5. Note that I (or A) is an indicator of the average intensity of poverty among the poor, but is not a good overall measure of poverty. In particular, it can increase when a poor person escapes poverty and the remaining incomes are unchanged, thus violating a standard monotonicity requirement.

6. Foster and Sen (1997) discussed the trade-off between the desirable properties of the new aggregate measures and the simplicity of the headcount measures and other "partial indices" of poverty that convey tangible information on one aspect of poverty.
7. Precise definitions can be found in Foster and Sen (1997), Foster (2005) and Foster et al (2013).
8. By linear replication it is meant that a replication of the distribution that results in a *k*-fold increase in population leads to a *k*-fold increase in measured poverty. Additivity is captured in equations (5) or (6) in the text.
9. It also indicates how well changes in H predict changes in H_e. Note that the growth elasticity of H_e (or equivalently, of P_1) is the product of the depth elasticity ε and the traditional growth elasticity of H.
10. In a similar fashion, the number of full time equivalent employees at a company can exceed the number of persons employed if average hours worked at a company exceeded a benchmark such as 40 hours per week.
11. This would not be true if the goal were to lower person equivalent poverty to a particular absolute level (or indeed by an absolute amount), since absolute levels and changes depend on the benchmark. In this case it would be crucial to specify the benchmark ahead of time.
12. PovCalNet was accessed July 2015 at http://iresearch.worldbank.org/PovcalNet/index.htm.
13. Note that data from the Latin America and Caribbean countries in PovCalNet use income data, whereas all other countries (except Latvia) use consumption data. A few countries in LAC have both consumption and income data available from the same year. For these countries poverty rates are substantially higher using income data, and the differences between income and consumption data are even greater for the poverty gap than they are for the conventional headcount. Given these differences between consumption and income data – which the person equivalent headcount measure highlights – comparisons between LAC and other regions should be interpreted with caution.
14. A multilevel development goal sets targets and evaluates outcomes for several levels of population aggregation.
15. Foster and Shorrocks (1991, p. 699) derives the class $P_\lambda = \lambda H + (1-\lambda)P$ axiomatically where P is a continuous, decomposable measure and $\lambda \in [0,1]$. A similar form arises in the measurement of ultra-poverty. See Foster and Smith (2015).
16. For example, the other major target of the first MDG is halving hunger. Hunger is generally expressed in terms of the fraction of the population projected to suffer from a below-minimum caloric intake. Yet surely people moving toward that minimum, even if not yet crossing it, also represents progress against hunger. The person equivalent approach could be a valuable complement to existing headcount metrics for international poverty goals.
17. See also Alkire et al. (2015).
18. See Alkire and Santos (2014).
19. When comparing H_e with H_e', for clarity we refer to the former as the depth p. e. headcount ratio, and the latter as the severity p. e. headcount ratio.

References

Alkire, S., J. Foster, M. Santos, S. Seth, P. Ballon and J. Roche (2015) *Multidimensional Poverty Measurement and Analysis*. Oxford: Oxford University Press.

Alkire, S., and Foster, J. (2011) "Counting and Multidimensional Poverty Measurement,"' *Journal of Public Economics,* vol. 95, nos 7–8, pp. 476–87.

Alkire, S., and Santos, M. (2014) "Measuring Acute Poverty in the Developing World: Robustness and Scope of the Multidimensional Poverty Index," *World Development*, vol. 59, pp. 251–74.

Atkinson, A. (1970) "On the Measurement of Inequality," *Journal of Economic Theory*, vol. 2, pp. 244–63.

Basu, K., and H.V. Pham (1998) "The Economics of Child Labor," *American Economic Review*, vol. 88, no. 2, pp. 412–27.

Bourguignon, F., and G. Fields (1990) "Poverty Measures and Anti-Poverty Policy," *Recherches Economique de Louvain*, vol. 56, nos 3–4, pp. 409–27.

Foster, J., J. Greer and E. Thorbecke (1984). "A Class of Decomposable Poverty Measures". *Econometrica*, vol. 3, no. 52, pp. 761–6.

Foster, J. (2005) "Poverty Indices," in A. de Janvry and R. Kanbur (eds), *Poverty, Inequality and Development: Essays in Honor of Erik Thorbecke*. Amsterdam: Kluwer Academic Publishers.

Foster, J., M. Lokshin, Z. Sajaia and S. Seth (2013) *A Unified Approach to Measuring Poverty and Inequality: Theory and Practice*. Washington, DC: World Bank Press.

Foster, J.E. and A.K. Sen (1997) "On Economic Inequality. After a Quarter Century," Annex to the enlarged edition of A'K. Sen, *On Economic Inequality.* Oxford: Clarendon Press.

Foster, J., and S. Smith (2015) "Measuring Ultrapoverty: A New Class of Indices and an Application to Rural Ethiopia Panel Data." Unpublished.

Foster, J., and A. Shorrocks (1991) "Subgroup Consistent Poverty Indices,"' *Econometrica,* vol. 59, no. 3, pp. 687–709.

Risley, H. (2012) "The Economic Perspective: Causes and Constraints to Reaching the Ultra Poor," USAID Microlinks Blog. https://www.microlinks.org/blogs/economic-perspective-causes-and-constraints-reaching-ultra-poor.

Sen, A. (1976) "Poverty: An Ordinal Approach to Measurement," *Econometrica,* vol. 44, no. 2, pp. 219–31.

Sen, A.K. (1992) *Inequality Reexamined*. Cambridge, MA: Harvard University Press.

United Nations (2014) *2014 Millennium Development Goals Report*. New York: United Nations.

Watts, H.W. (1968) "An Economic Definition of Poverty," in D. P. Moynihan (ed.), *On Understanding Poverty*. New York: Basic Books.

World Bank (2010) *The MDGs after the Crisis: Global Monitoring Report 2010*. Washington, DC: World Bank.

World Bank (2013) *The State of the Poor: Where are the Poor and where are they Poorest?* http://www.worldbank.org/content/dam/Worldbank/document/State_of_the_poor_paper_April17.pdf.

Comments on "Person Equivalent Headcount Measures of Poverty" by Tony Castleman, James Foster, and Stephen C. Smith

Bhaskar Dutta
Department of Economics, University of Warwick

Introduction

Castleman, Foster and Smith (henceforth CFS) start from the observation that while the simplicity of the poverty headcount measure has made it a very popular measure particularly amongst policy makers, it has well-known shortcomings. In particular, the headcount measure, through its single-minded focus on the *numbers* of people below the poverty line, ignores completely the *intensity* of poverty. CFS contend that measures which are more satisfactory from a conceptual perspective are also more complicated and hence less comprehensible to most policy makers.

This prompts CFS to attempt the construction of a new measure which retains the essential simplicity of the headcount measure while at the same time avoiding its shortcomings. They define a new and interesting measure – the *Person Equivalent Headcount Measure* (PEHM) – and show how it can be applied to the unidimensional context, that is of *monetary* poverty. In order to illustrate the measure, the paper also calculates person-equivalent headcounts for various countries and regions given a monetary poverty line of $1.25 a day.

In this note, I briefly discuss the extent to which the PEHM is consistent with the main motivation underlying the paper.

The person equivalent headcount measure

Let n be the number of individuals. Throughout the note, the population size is fixed for simplicity – CFS also discuss how their approach can be adapted to the case of variable population.

The first step in the measurement of poverty is identifying the set of poor persons. Of course, the identification exercise in the unidimensional case is straightforward. It involves specifying a poverty line z and identifying a person

as poor if $y_i < z$. Let $q(y)$ be the number of people who are poor corresponding to the distribution y. Then, the traditional headcount measure is

$$H = q(y)/n$$

A poor person *i*'s *intensity* of poverty is given by $s_i = z - y_i$. So, the average intensity of poverty is given by

$$A(y) = \sum_i s_i / q$$

where, of course, the sum is taken over all persons who are poor. A major shortcoming with the headcount measure is that it will declare two distributions y and y' to have the same level of poverty so long as $q(y) = q(y')$, even if for instance $A(y)$ is significantly higher than $A(y')$.

CFS advocate the following procedure. Choose a benchmark or base year 0 and the corresponding A^0. Their *person-equivalent headcount measure* (PEHM), q_e, is the number of poor persons with the initial average shortfall A^0 that it would take to aggregate up to the new total shortfall of $A^1 q^1$. That is

$$qe = A^1 q^1 / A^0$$

Since $I^1/I^0 = A^1/A^0$, q_e can be written alternatively as

$$q_e = q^1 \, I^1/I^0$$

CFS call the ratio A^1/A^0 (or equivalently I^1/I^0) the *transformation factor*. So, the PEHM is obtained by multiplying the actual number of the poor by the transformation factor. Multiplication of the traditional headcount by the transformation function therefore counts the number of person equivalents.

The first issue is whether to view q_e as a measure of the *level* of poverty in any given year. Since the main motivation of the paper is to find "simple" measures that appropriately reflect the intensity of poverty, CFS presumably do want to interpret PEHM as a measure of the level of poverty.[1] Unfortunately, measuring the *level* of poverty in year 1 with reference to a base year has an obvious problem – the incidence of poverty in any year now depends on the benchmark or base year value of A^0. Change A^0 and the level of poverty in year t will change even when there is no change in the year t income distribution – this is surely somewhat counter-intuitive.

It is also important to be able to measure the change in the incidence of poverty across time. Given a measure of poverty P, the obvious measure of change is $P^2 - P^0$,[1] where the superscripts refer to two time periods. CFS mention that the traditional way of "assessing progress is by using the change in headcounts".

They dismiss this method since it ignores any possible change in intensity of poverty between the two time periods. Having defined qe, CFS define positive progress or an improvement if $qe - q^0 < 0$, while a worsening in the incidence of poverty would be indicated if $qe - q^0 > 0$.

This is more satisfactory than the difference in headcounts since it also incorporates the change in intensities. While this is an interesting way of measuring change, it is not clear that it is a particularly *transparent* or *simple* method of measuring change.

Let us consider other options. Define $g_i = s_i/z$ if i is poor and $g_i = 0$ otherwise. Then, one measure which does take into account the intensity of poverty is the average normalized income shortfall or gap from the poverty line.

$$P_1 = \mu(g_1, \ldots, g_n)$$

So, P_1 is the sum of the income gaps of the poor divided by the size of the population. Change in the incidence of poverty can be measured as the difference in average normalized income gaps –

$$\Delta P_1 = P_1^1 - P_1^0$$

The concept of an average is simple enough. Nevertheless, CFS feel that even the average normalized poverty gap may be too complicated for policy makers. However, CFS show that

$$q_e = nP_1/I^0$$

This should surely imply that the PEHM should be more difficult for policy makers to grasp than P^1! After all, one more algebraic operation is required to compute the PEHM after P_1 has been computed.

Notes

1. Indeed, one of the authors confirms this interpretation in private correspondence.
2. Alternatively, one could use P^1/P^0.

4

How Useful Is Inequality of Opportunity as a Policy Construct?

Ravi Kanbur
Cornell University

Adam Wagstaff
World Bank

4.1 Introduction

In policy and political discourse, "equality of opportunity" is the new motherhood and apple pie. It is often contrasted with equality of outcomes, with the latter coming off worse. Equality of outcomes is seen variously as Utopian, as infeasible, as detrimental to incentives, and even as inequitable if outcomes are the result of differing efforts. Equality of opportunity, on the other hand, is interchangeable with phrases such as 'leveling the playing field', 'giving everybody an equal start' and 'making the most of inherent talents.' In its strongest form, the position is that equality of outcomes should be irrelevant to policy; what matters is equality of opportunity.

But the application of the broad concept of equality of opportunity requires that it be specified sufficiently precisely to be measurable and implementable. The last two decades have seen a number of attempts to formulate the concept of equality of opportunity with a view to policy application, not least in the World Bank through its *World Development Report* 2006 (World Bank, 2005) and the subsequent "mini-industry" which has developed in estimating measures of inequality of opportunity using household survey data (for example, de Barros et al. 2009). The empirical work appears to have given confidence to those who would use the concept of equality of opportunity in the policy context, that it can be measured and quantified, so that policies can be designed and judged according to it.

However, attempts to quantify and apply the concept of equality of opportunity in a policy context have also revealed a host of problems of a conceptual and empirical nature, problems which may in the end even question the practical usefulness of the concept. The object of this paper is to review and critique recent attempts at specifying and quantifying equality of opportunity, and to assess the extent to which they do in fact allows us to discard equality of outcomes as a key part of policy evaluation.

The plan of the paper is as follows. Section 4.2 sets out the philosophical basics of equality of opportunity, focusing on key writings in the last three decades. Section 4.3 then introduces the attempts at application and quantification at a general level, particularly in the policy context, and presents a critique. Section 4.4 carries the argument forward to the specifics of the education and health domains, to see how the general arguments play out in these settings. Section 4.5 concludes.

4.2 Equality of opportunity: the basics

The core premise of the literature on equality of opportunity is the idea that inequalities in outcomes can be partitioned into justifiable (or legitimate) sources of inequality (let us call them the J's) and unjustifiable (or illegitimate) sources (let us call them the U's). The idea is that inequalities in outcomes are to be seen as just (or fair) insofar as they are driven by inequalities in the J's, and unjust (or unfair) insofar as they are driven by inequalities in the U's. The practical usefulness of the idea of inequality of opportunity boils down to two sets of questions: Is there agreement on the set of influences on the outcome in question and on whether each is a J or a U? And, conditional on a consensus or close to a consensus on these questions, can the inequalities be partitioned empirically in a fashion that is credible and meaningful for policy?

On the face of it, these are two separate exercises – the first conceptual (and normative), and the second empirical. But in practice many of the challenges at the conceptual level only surface when the empirical operationalization begins, and when researchers begin to grapple with a practical partition of influences on outcomes into the J's and the U's. As a result, while appearing to resolve the key conceptual questions, many of the philosophical discussions in the literature actually leave many key questions unanswered. Thus while this section, which focuses on the philosophical underpinnings of equality of opportunity, will highlight *some* of the conceptual issues involved, it will not resolve them all. Only after we have gone through the empirical applications in the next two sections will the full gamut of conceptual difficulties become apparent.

The modern flowering of the equality of opportunity literature in economics owes much to the work of Roemer (1998), who is recognized for example by the authors of the *World Development Report 2006* (see World Bank, 2005; Bourguignon, Ferreira and Walton, 2007). Roemer sets out ideas on both questions. The central conceptual distinction in Roemer is to "separate the influences on the outcome a person experiences into *circumstances* and *effort*: the former are attributes of a person's environment for which he should not be held responsible, and effort is the choice variable for which he should be held responsible" (Roemer, 2008; original emphasis). It is this distinction which

Roemer, and followers like de Barros et al. (2009), attempt to implement in empirical policy oriented work.

Roemer (1998) also suggests a way to partition inequalities empirically into justifiable inequalities (in his book, inequalities in effort) and unjustifiable inequalities (in his book, inequalities in circumstances), from which he arrives at a measure of inequality of opportunity. He addresses the question of how exactly to calibrate the effort of an individual, and argues that the rank of an individual in the effort distribution across individuals with the same circumstances should be the relevant metric. De Barros et al. (2009) use alternative metrics which we will discuss below. But the applications all rely on the assumption that circumstances and effort can indeed be separated. While the concepts are clear in principle, as we shall see the applications invariably raise the question of whether they can ever be separated in practice, or even conceptually. And as we will also see, questions arise as to whether there is more to the J vs. U distinction than 'circumstances' vs. 'effort.'

As Roemer (1998, 2008) recognizes, his contribution feeds into an earlier philosophical debate which was ignited by Ronald Dworkin (1981a, 1981b) when he asked the question "What is equality?" Dworkin argued essentially that certain types of preferences should not elicit an egalitarian redistributive response. For example, if a person would be deeply unhappy if unable to drive fast cars, this should not arouse an egalitarian sentiment to transfer resources to this person. On the other hand, if a person was born with low inherent abilities to prosper in the market place, this might be a legitimate basis for redistribution.

Of course, a lot, in fact everything, rests on coming to a separation of legitimate and illegitimate differences for the egalitarian impulse. As Cohen (1989) pointed out, what if preferences (even for fast cars) were the result of an upbringing which inculcated those preferences, or brooked no dissent from them? Or if preferences between work and leisure ("laziness") were inculcated in the same way? Thus a conceptual separation of influences into those which are legitimate to take into account in addressing inequality and those which are not merely pushes the issue to the next level of how this separation is to be made, conceptually and practically.

The above discussion can be related to a parallel discussion in economics on whether the distribution of ex ante or ex post utility should be the relevant object in social welfare assessment. Milton Friedman (1962) made the ex ante case:

> Consider a group of individuals who initially have equal endowments and who agree voluntarily to enter a lottery with very unequal prizes. The resultant inequality of income is surely required to permit the individuals in question to make the most of their initial equality… Much of the inequality

> of income produced by payment in accordance with the product reflects 'equalizing' difference or the satisfaction of men's tastes for uncertainty… Redistribution of income after the event is equivalent to denying them the opportunity to enter the lottery. (Friedman, 1962)

Further, Friedman is clear that this is not just a fanciful special example but that it captures strong elements of reality:

> This case is far more important in practice than would appear by taking the case of a 'lottery' literally. Individuals choose occupations, investments and the like partly in accordance with their tastes for uncertainty. The girl who tries to become a movie actress rather than a civil servant is deliberately choosing to enter a lottery, so is the individual who invests in penny uranium stocks rather than government bonds.

The argument, which clearly draws on Friedman's earlier work on choice under risk (Friedman and Savage 1948; Friedman 1953), encapsulates the central point that the observed inequality of income may overstate ex ante inequality. In the pure case, there is no ex ante inequality at all and yet risk produces ex post inequality. But for Friedman this ex post inequality has no claims on our moral intuitions and should not elicit a redistributive impulse – to do so otherwise would be "equivalent to denying them the opportunity to enter the lottery."

In the more general case, when individuals differ in endowments, observed inequality will be a combination of the two effects. But Friedman's pure case is nevertheless useful for smoking out core intuitions on whether outcomes or opportunities matter more. Suppose in the pure case one of the outcomes is destitution – below a poverty line, say. Recall that by construction there is perfect equality before the lottery. Further, all individuals freely choose to participate in the lottery. To use the terminology of Dworkin (1981b), this is "option luck", chosen by the individual, not "brute luck" over which the individual has no control. And yet, upon exercising this free choice starting with perfect equality, after the lottery is drawn not only is there inequality, but some are destitute. The outcome is brutal even though the luck was not brute. When faced with such destitution, do our moral intuitions guide us to redistribute after the event, or do they guide us to stand firm because to redistribute would be in effect to deny the consequences of equality of opportunity (and the consequences of the doctrine of responsibility for actions as advanced by Dworkin)?

Kanbur (1987) has argued that in the case of destitution as an outcome of the lottery choice, our moral intuitions do indeed veer towards ex post redistribution and support for the destitute. To emphasize the point, imagine yourself serving on a soup line of the indigent. Consider then the idea that we would

condition the doling out of soup on an assessment of whether it was circumstance or effort which led to the outcome of the individual in front of us to be in the soup line. Surely this is morally repugnant, and it establishes that at least for extreme outcomes the outcome-based perspective dominates any considerations of opportunity. Indeed, this point is taken on board by Bourguignon, Ferreira and Walton (2007) in their definition of equitable development policy which "makes avoidance of severe deprivation a constraint that must be satisfied in the process of pursuing the broader objective of equal opportunity."

Notice that none of the above is to deny that outcome-based redistribution will have incentive effects. Indeed, in general so will an opportunity-based intervention, or any intervention. In the usual way, as in the classic work of Mirrlees (1971), the incentive effects will have to be balanced against the egalitarianism of the outcome-based social welfare function. The point, rather, is that in a setting designed to strengthen the case for the opportunity-based perspective to the utmost, outcome-based concerns nevertheless come through, and ex post inequality of outcomes enters the evaluation even when there is perfect ex ante equality of opportunity.

The Friedman example can be used to elaborate on a conceptual difficulty with the concept of equality of opportunity already alluded to. If parents engage in the Friedman lottery, the outcomes will lead to inequality among their children even though, by construction, there is perfect equality of circumstance among the parents. Parents who win big in the lottery will be able to spend more on their children's education, for example, than parents who lose big from the decisions they have all made with full knowledge of the outcomes. The lottery outcomes are now the circumstances of the children. The doctrine of equality of opportunity would now lead us to redistribute in order to address inequality of circumstance among children. But, equally, it would lead us not to redistribute across parents because to do so would be "equivalent to denying them the opportunity to enter the lottery." Equality of opportunity, it would seem, is caught between two inconsistent Old Testament Biblical injunctions from Deuteronomy: "for I, the Lord your God, am a jealous God, visiting the iniquity of the fathers on the children, and on the third and the fourth generations of those who hate Me" versus "Fathers shall not be put to death for their sons, nor shall sons be put to death for their fathers; everyone shall be put to death for his own sin."

Of course, the above example with parents and children takes us back to other examples in Cohen (1989), which highlighted difficulties in Dworkin's (1981a, 1981b) attempts to introduce the notion of responsibility into egalitarian philosophy by drawing a distinction between preferences and resources. If preferences are themselves determined by resources, say parental resources, then a clean separation may not be possible, certainly empirically and perhaps even conceptually. These difficulties are the subject of a large literature and

debate. Key perspectives in this, from Roemer (1998), Barry (2005), and Swift (2005), are summarized by Jusot, Tubeuf and Trannoy (2013) as follows:

> In what we call Barry's view, circumstances are past variables and efforts are the variables which reflect the free will of the present generation. In Roemer's view, the vector of circumstances includes all past variables, and the descendant's efforts must be cleaned from any contamination coming from circumstances. In Swift's view, the vector of circumstances only includes past variables, which have no consequences on children's efforts. In other terms, the vector of circumstances must be cleaned from any correlation with [children's] efforts. (p. 1473)

The differences between these views are only one part of a large literature on a range of issues in equality of opportunity[1] which reflect basic ethical disagreements that are unlikely to be resolved easily, and in any case need to be supplemented by the case of extreme outcomes discussed above, and how our moral intuitions in such extreme cases translate in turn to less extreme conditions.

4.3 Implementation in a policy context: income inequality

The ethical disagreements on basic concepts notwithstanding, implementation of measures of (in)equality of opportunity has proceeded apace. The burgeoning literature has provided a plethora of measures and applications to data sets, from early attempts by Roemer and his associates (for example, Roemer et al. 2003), to more recent applications in specific contexts like health (for example, Jusot, Tubeuf and Trannoy, 2013). Thus the partitioning of influences on outcomes into legitimate and illegitimate has now been undertaken empirically in a large number of contributions across many countries.

Of particular interest is a line of analysis established by the work of de Barros et al. (2009). In part this is because it has become popular throughout the World Bank (spreading from Latin America, where it was developed originally, to other Bank regions), and through that channel to the broader policy-oriented literature. Its popularity no doubt owes much to its simplicity, being relatively straightforward to apply with readily available household survey data. In part, though, our interest in the work stems from the fact that in empirically operationalizing the equality of opportunity approach researchers are forced to confront a whole host of tough questions that are more easily overlooked in conceptual discussions. How satisfactorily these questions can be answered at the empirical level determines ultimately the practical usefulness of the whole equality of opportunity approach.

There are, in fact, two somewhat separate approaches in de Barros et al. (2009) and we will take these up in turn in this section and in the next. In the

first approach the primary outcome variable is the usual variable used to calculate standard inequality and poverty measures – earnings, income or consumption. The conceptual basis and implementation method is stated succinctly by de Barros et al. (2009) as follows:

> To measure inequality of opportunity for a certain outcome, total inequality in the outcome can be decomposed into two parts: one resulting from circumstances beyond individual control and a second part resulting from unequal individual effort and luck… First, six variables related to circumstances exogenous to the individual were identified from the most comprehensive data sets available: gender, race or ethnicity, birthplace, the educational attainment of the mother, the educational attainment of the father, and the main occupation of the father… Then the sample was partitioned (in each country) into groups or "cells," such that all individuals in any given cell have exactly the same combination of circumstances. The resulting subgroups are known in the literature as "types." These cells are then compared with one another. The difference in outcomes between cells can be attributed to inequality of opportunity, while the differences within cells can be considered the result of effort or luck." (pp. 125–6)

A decomposable measure of inequality, the Theil-L index, is used in a non-parametric method of quantifying the extent of variation in income accounted for by within cell variation and between cell variation, the latter the being identified as "inequality of opportunity" (IEO).

It is easy to see why this method has caught on. It is relatively simple and intuitive. Indeed, it is nothing more than standard inequality decomposition analysis which has been conducted in economics and development economics for a long time (Kanbur, 2006). For example, the percentage of inequality accounted for by caste has been a staple of the Indian discourse, as has the contribution of ethnic differences to inequality in Malaysia, as has the north–south divide in Ghana, and so on. What is new is twofold. First is its systematic application in a consistent manner across a range of countries as done for Latin America in de Barros et al. (2009). Second, crucially, is the interpretation and labeling of the between group component in the inequality decomposition as inequality of opportunity, thereby linking to the philosophical literature discussed the last section, and also thereby linking to the appeal in the policy discourse of "equality of opportunity" as opposed to equality of outcome.

On the basis of this method, de Barros et al. (2009) conclude that "between a quarter (Colombia) and half (Guatemala) of the income inequality that we observe among adults in Latin America is due to the circumstances they faced when they started out in life – at the very outset, through no fault of their own. And while their race, sex, and location all played a role, no circumstances

were more powerful than their mothers' education and their fathers' incomes" (p. xviii). Again, that race, sex, location, and parental education and income are determinants of outcomes in the income dimension is not a surprise – it is a commonplace finding in almost every non-parametric and parametric analysis of income distribution. What is new is the additional interpretation that between 25% and 50% of outcome inequality can be attributed to circumstance variables and therefore to inequality of opportunity.

How might such findings impact on the policy debate? When presented with these findings, how might policy makers react? One response might be to say "You mean inequality of opportunity is as high as 50%? Gosh, we must do something about it!" But equally another response might be "You mean inequality of opportunity is only 25%? Gosh, we are doing quite well on the equity front so let us focus on other priorities!" Indeed, the second response might even be made if the figure was 50% rather than 25%.

The response of de Barros et al. (2009) to the possibility that such quantification might lead to a downplaying of the need for redistribution is to emphasize the fact that these estimates are in fact a lower bound on inequality of opportunity:

> …although the six variables employed in this chapter are a richer set of circumstances than those used in any previous study known to the authors, it is still possible to think of other relevant variables that are not observed. A "true" measure of inequality of opportunity would require using all relevant circumstance variables to partition the population into types. But this is, of course, extremely unlikely to be feasible in practice for any conceivable data set, and it is certainly impossible for the seven countries examined here. The empirical estimates defined in this chapter – regardless of whether parametric – should therefore be interpreted as lower-bound estimates of inequality of opportunity; including any additional circumstances would cause each cell to be further subdivided. (p. 127)

The difficulty, of course, is that the lower bound is just that. An alternative number which exceeds the estimate is equally valid. Thus in the case where the contribution of the circumstance variables is estimated as 25%, the contribution could equally well be 75%. But the fact that 25% is the number that is produced in front of the policy makers will make it akin to a point estimate in the policy discourse, no matter how much the analyst caveats it as a lower bound. Thus an estimate of the IEO is not quite the same as other standard poverty or inequality measures. We cannot say how far the IEO of a specific country at a specific date falls short of its true value. Countries could presumably have the same true value but have different estimated values, and vice versa. A country's estimated IEO value could presumably fall over time without its true value

falling. This makes an IEO estimate potentially highly misleading. And it marks out the IEO as different from a poverty or inequality number. There may be data issues that make poverty and inequality estimates problematic, but they do not lead us to systematically underestimate what we are trying to measure by an inherently unknown amount.

To this must be added another worrying aspect of the empirical approach. The extent of inequality of opportunity reported to policy makers will be dependent on whether or not the data set collects certain types of information. So estimates across countries will either be determined by the lowest common denominator of commonly available variables, or will be simply not comparable, even as a lower bound.

The reservations above all take for granted de Barros et al.'s (2009) partitioning of influences on outcomes into illegitimate and legitimate. In the former, they include gender, race or ethnicity, birthplace, the educational attainment of the mother, the educational attainment of the father, and the main occupation of the father. All other influences necessarily fall into the group of legitimate influences. De Barros et al. (2009) explicitly highlight – with approval – the inclusion of effort and luck among these. Yet if some of the variation is due to "brute luck", in the words of Dworkin (1981b), then much of the philosophical discourse would argue for it to be a legitimate cause for intervention. Indeed, the previous section has argued that even for "option luck", which is the result of individual choices, the extreme low end of variation in outcomes is also a legitimate cause for intervention. These considerations are swept under the carpet when inequality decompositions are presented as estimates of "inequality of opportunity."

4.4 Children, education and health

Questions about the practical usefulness of the concept of equality of opportunity become even tougher when the focus shifts from income and consumption to education and health, especially when the exercise involves children, who, even in health, end up hogging the spotlight in such exercises.

The de Barros et al. (2009) study is again a useful place to start. This is in part because of its influence, but in part because it illustrates nicely the challenges faced by anyone wanting to make the case that the equality of opportunity approach has practical usefulness. While income is the main focal variable in the IEO work, the same methodology is also used to partition inequalities in educational attainment among 15-year-olds into legitimate and illegitimate sources of inequality. The same concerns about the methodology – notably the estimates being a lower bound – apply with equal force to educational attainment as to income. Conceptual and normative questions to do with the partitioning of influences into the J's and the U's also apply. But these also

apply to the second methodological approach developed by de Barros et al. – the Human Opportunities Index (HOI) – to which we now turn before dealing with the tough normative questions that arise in the context of education and health in both exercises.

In many ways, the HOI is the authors' preferred approach in de Barros et al. (2009), because it focuses on children and on access to basic services, including education and health. The strong suggestion in the motivating discussion of the study is that with this focus there is likely to be a strong consensus on the policy objectives:

> This study defines basic opportunities as a subset of goods and services for children, such as access to education, to safe water, or to vaccinations, that are critical in determining opportunity for economic advancement in life. These are either affordable by society at large already, or could be in the near future, given the available technology. Universal provision of basic opportunities is a valid and realistic social goal. In the case of children, most societies agree on the importance of a set of basic opportunities, at least at the level of intentions; even if different societies might have different standards about the right set of basic opportunities, there is some global consensus on a few of them, just as there is consensus regarding the Millennium Development Goals. Here we include as basic opportunities variables related to education (completion of sixth grade on time, and school attendance at ages 10–14) and housing conditions (access to clean water, sanitation, and electricity). (p. 3)

Thus the authors start with the outcome variable as being access to social infrastructure such as education, health, water, electricity, and so on. The empirical focus is on the access of children to these services. Each child either does or does not have access to the service in question. The authors first estimate how the probability of a child's access depends on the child's circumstance variables: "parents' education, family per capita income, gender, family structure (number of siblings, single-parent household) and area of residence (urban versus rural)" (p. 66). The predicted probability clearly varies with the circumstances. This predicted probability is then compared to the average probability of access across all circumstance variables. The differences in the predicted probability relative to the average are then used to construct a measure of inequality in access, and this is the inequality of opportunity. The specific measure used is a form of dissimilarity index based on the absolute difference between each individual value and the average. The authors further argue that the average access measure can be deflated by the dissimilarity index to give a measure of effective access correcting for inequality of opportunity – their Human Opportunity Index (HOI).

A number of specific questions can be raised on the dissimilarity index and the implicit weights used for small and large departures from the average. Alternative weights would lead, of course, to different measures of inequality of opportunity as formulated. But a more important point is that analogous criticisms can be made to this approach as were applied to the first approach discussed in section 4.3. In the limit, if enough variables are used, each child's access can be predicted perfectly so we are measuring inequality of the binary variable access/no access. Using fewer variables than this will reduce the dissimilarity index. Put another way, adding more circumstance variables will increase the measured inequality of opportunity. Thus, in this case as well, the number put out for policy purposes is a lower bound and open to similar notes of caution in the policy discourse.

In fact, as one of us has argued elsewhere (Wagstaff, 2013), the whole empirical approach used by de Barros et al. (2009) in their HOI is misplaced, given what they say early on in the book about the appropriate partitioning of influences into J's and U's:

> … from an empirical standpoint, the principle of equality of opportunity as "leveling the playing field" can be readily operationalized by measuring children's access to basic goods and services that are critical for the full development of a child. For children, access defines "opportunity," because children (unlike adults) cannot be expected to make the efforts needed to access these basic goods by themselves. (p. 3)

Thus despite apparently taking the view that *all* inequality in these childhood variables is unjust, de Barros et al. (2009) develop a measure of inequality of opportunity that picks up only *a part* of the observed inequality. There is an inconsistency here that is puzzling to say the least.

This raises the more general question of what influences on education and health variables belong in the J's and what belong in the U's. The partitioning of certain influences on education and health outcomes – notably parental influences, luck and talents – has proved especially challenging. It is worth going through these issues carefully.

4.4.1 One person's outcomes depend on another person's efforts

Most of the outcome indicators investigated by de Barros et al. in HOI relate to children. In early childhood, it is the effort of the parent – not that of the child – that shapes outcomes. A one-year-old infant cannot be said to exert effort to be immunized (one of the outcomes examined by de Barros et al. (2009)); rather, it is the parent who makes or does not make the effort. The same is true of other outcomes in infancy. At the very minimum then it would seem that all inequalities among infants should be deemed unjust. This is, in

fact, the stated view of de Barros et al. (2009), even though in their operationalization of the HOI they deviate from this stance.

As a child moves through infancy into childhood then into the teenage years and then to the cusp of adulthood, the role of the child's own effort in shaping outcomes increases. This is truer of individual-level outcomes than of household-level outcomes, such as whether a family has access to safe water – another of the outcomes considered by de Barros et al. (2009). These household-level outcomes reflect almost entirely the decisions and efforts of parents, and it seems likely that most people would agree that inequalities among children in water and sanitation are unjust no matter how they arose. Again this is the stated view of de Barros et al. (2009) but not the view implied by their decomposition.

What of inequalities in individual-level outcomes after infancy? Take inequalities in primary school completion and educational attainment at age 15 – two of the other indicators used by de Barros et al. (2009). Insofar as these reflect inequalities in children's efforts and choices, are these inequalities unjust? De Barros et al. (2009) argue that inequalities in primary completion reflect parental effort and are hence unjust, even though – as with the other childhood indicators – they go on to separate out the part of the inequality due to inequalities in circumstance. By contrast, in their IEO exercise they group inequalities in educational attainment at age 15 with inequalities in income, and in both cases strip out the contribution from inequality of circumstance; the rationale in both cases is that the part of the inequality due to inequality in effort is not unjust.

Roemer (1998) – whose work inspired de Barros et al. (2009) – has a different viewpoint. He argues that insofar as they reflect inequalities in parental pressure or influence, inequalities in childhood outcomes (including presumably inequalities among young teenagers) should be counted as inequalities in circumstance, not as inequalities in effort; Roemer would want to extract from the inequality in outcome both the part caused by inequality in circumstance and the part caused by inequality in parental pressure. De Barros et al.'s de facto position is actually closer to that of Barry (2005), who argues that the inequality in parental effort and influence should not be parceled out, and that it is a just source of inequality.

The argument of de Barros et al. (2009) could be that we are constrained by data which do not allow us to separate out the effects of parental influence, and that their estimate of HOI, just like their estimate of IEO, contains an underestimate of the true inequality of opportunity. But we are then up against the same set of issues as discussed in the previous section – the danger that the value of HOI is taken not as an underestimate but as a point estimate; and that the value of HOI across countries may be determined simply by differences in data availability, not true differences in equality of opportunity.

4.4.2 Luck and risk

De Barros et al. (2009) argue that inequality in outcomes due to differences in luck is just: "In an ideal world, inequality in outcomes should reflect only differences in effort and choices individuals make, as well as luck" (p. 15). This has echoes of Friedman's (1962) lottery example, and prompts several questions.

Does it matter whether the person freely takes a risky decision? What about children whose educational attainment and health depend on the choices their parents make as well as their luck? If a child is unlucky enough to contract a respiratory disease by inhaling the smoke from his parents' cigarettes, is that still just? If a child performs badly in his school exams at age 15 because he is unlucky enough to see his parents separate and then divorce in the year before his exams, should we take this bad luck into account in assessing the justness of differences in educational attainment? Note this is not quite the same as the case of parental effort. Not every passive smoker falls ill. And some marriages keep going while others end, despite the best efforts of the two parties to keep it together. In both cases, the child got unlucky, but in neither case did the child freely take a risky decision.

What if the behavior was shaped by parental influences during childhood? Suppose, for example, a child acquires a taste for smoking or excessive drinking during youth by living in a home where smoking and drinking are the norm. This relates back to the debates between Dworkin, Roemer and Swift reviewed in section 2. Roemer at least would want these influences removed.

Do we think differently about luck depending on whether the risky activity is essential to a person's flourishing as a human being? Nobody needs to smoke or consume alcohol excessively, for example; in effect, the default choice is not to engage in these activities. People take a proactive decision to deviate from the default in the pursuit of short-term pleasure knowing they raise their risk of illness and premature death. By contrast, people do need to eat; eating is the default choice. Moreover, many would argue that for various reasons – including convenience, cost, and commercial pressures – the default diet today is a diet that poses risks to health, and that people have to make a conscious and determined effort to eat in a way that lowers health risks. This is not a trivial comparison: dietary risks are estimated to have accounted for more deaths worldwide in 2010 than alcohol and tobacco combined (Lim et al., 2012).

Does it matter how much is known about the risks involved and how well publicized they are? The risks associated with smoking and excessive alcohol consumption are well known and well publicized, but the evidence on diet is more complex, more fluid, and less well publicized. For example, given the attention they receive in the media, one might imagine the big culprits in relation to diet are too little polyunsaturated fatty acid, and too much processed meat, trans-fatty acids, sugar-sweetened beverages and red meat. Yet these are not actually the biggest causes of diet-related deaths worldwide: over six times

as many deaths are attributable to people consuming too much sodium, and too little fruit, nuts, seeds, vegetables and whole grains (Lim et al., 2012).

These issues aside, should we hold people accountable for bad luck, or only for risky behavior? De Barros et al. (2009) and Friedman (1962) argue the former. But there is an alternative school of thought that argues the opposite (see, e.g., Le Grand 1987; Cappelen and Norheim 2005), namely that holding people accountable for outcomes is too strong, and that people should not be held accountable for their bad luck ("brute luck," as Dworkin calls it); in this view, luck is just one of the many of Roemer's "circumstances beyond a person's control." Insofar as is feasible, smokers and drinkers should, according to this viewpoint, receive whatever additional health care is required to reduce their odds of premature death to the odds faced by everyone else. The accountability for risky behavior ("option luck" as Dworkin calls it) comes in through taxation: consumption of tobacco and alcohol (as well as, e.g., sodium, red meat, sugar-sweetened beverages, etc.) should be taxed at such a rate that the revenues cover the extra expected health care costs.

4.4.3 Talent

De Barros et al. (2009) argue that inequality in outcomes stemming from differences in talents is just: "Success in life should depend on people's choices, effort, and talents, not on their circumstances at birth" (p. 1). This is also a contestable position. They are not alone in taking this stance, but it is a contestable one.

The *Oxford English Dictionary* defines talent as a "power or ability of mind or body viewed as something divinely entrusted to a person for use and improvement". Thus people are endowed with a natural or innate talent, and this talent can be both used and cultivated. A talent cannot be cultivated from nothing, so in this sense endowed talents impose limits on the talents people can have in later life. One can become wealthy without inheriting wealth, but one would find it hard to say the least to become a concert pianist without some endowment of musical talent.

Since endowed talents are by definition beyond an individual's control, it is odd that de Barros et al. (and others) are so quick to accept as just inequalities stemming from inequalities in talents. More in keeping with Roemer's approach would be an attempt to parcel out the part of the inequality in outcome stemming from inequality in endowed talent and treat this as a source of inequity; the remaining part – due to differences in acquired talent – would be considered equitable. One might argue that one should go further and strip out some of the inequality due to inequality in acquired talent. After all the degree to which people can cultivate an innate talent depends on their circumstances during childhood and on efforts made by parents; inequality arising from the first is agreed to be unjust by everyone, and for those like Roemer inequality stemming from the second is also unjust.

4.5 Conclusion

None of the above is to question the enormous contribution on the determinants of inequality of outcomes that has been made by the empirical literature which tries to measure and quantify "inequality of opportunity" in an attempt to make the concept more policy-relevant. De Barros et al. (2009) and the related literature have in effect analyzed the determinants of a range of outcome variables. Such analyses have always been conducted, but the rubric of "equality of opportunity" has given a push to this analysis in the empirical domain. And if the use of the label "equality of opportunity" opens doors with policy makers to present results to them which would otherwise be ignored if they were labeled "equality of outcomes", or simply "equality" or "equity", then this development is to be welcomed for that reason as well.

But the use of the aura of equality of opportunity as a concept, as a metaphor and as a label comes with its own problems, and these problems are magnified when the concept is applied and implemented with a view to contributing to the policy discourse on inequality. Health inequality is emblematic of the difficulties that current approaches face. If children's health is truly outside their control, then all of the inequality in their health is a legitimate objective of policy, not just that part which is explained by variables which measure parental circumstances. Similarly, especially for children but also for adults, if bad luck leads to ill health then wiping out this inequality as illegitimate for policy concern does not sit well with moral intuition – and yet that is what the present procedures which calculate inequality of opportunity in health tend to do.

At the same time, the present exercises skirt some fundamental questions in equity of outcomes (cf., e.g., Culyer and Wagstaff, 1993). One concerns the evaluation of the joint distribution of income and health. For adults, should health inequality be assessed as a stand-alone phenomenon, or should the correlation between income and health be a key normative criterion over and above the inequality of health and income taken separately (cf., e.g., Atkinson, 2011; Bleichrodt and van Doorslaer 2006; Fleurbaey and Schokkaert, 2011)? Such specific normative questions, and the specific causal determinants of different outcomes, are more likely to be useful to the policy discourse than general overarching attempts to quantify an abstract notion of "equality of opportunity."

Note

1. A recent selection would include Fleurbaey and Peragine (2013), Ooghe, Schokkaert and Van de Gae (2007) and Peragine (2004).

References

Atkinson, Anthony B. (2011) "On Lateral Thinking," *Journal of Economic Inequality*, vol. 9, no. 3, pp. 319–28.

Barry, B. (2005) *Why Social Justice Matters*. Cambridge: Polity Press.

Bleichrodt, Han and Eddy van Doorslaer (2006) "A Welfare Economics Foundation for Health Inequality Measurement," *Journal of Health Economics*, vol. 25, no. 5: pp. 945–57.

Bourguignon, Francois, Francisco Ferreira and Michael Walton (2007) "Equity, Efficiency and Inequality Traps: A Research Agenda," *Journal of Economic Inequality*, vol. 5, pp. 235–56.

Cappelen, A.W. and O.F. Norheim (2005) "Responsibility in Health Care: a Liberal Egalitarian Approach," *Journal of Medical Ethics*, vol. 31, no. 8, pp. 476–80.

Cohen, G.A. (1989) "On the currency of egalitarian justice," *Ethics* 99, 906–44.

Culyer, Anthony J. and Adam Wagstaff (1993) "Equity and Equality in Health and Health Care," *Journal of Health Economics*, vol. 12, no. 4, pp. 431–57.

de Barros, Ricardo, Francisco H.G. Ferreira, Jose R. Molinas Vega and Jaine Saavedra Chanduvi (2009) *Measuring Inequality of Opportunities in Latin America and the Caribbean*. Washington, DC: The World Bank.

Dworkin, R. (1981a) "What is Equality? Part One: Equality of Welfare," *Philosophy and Public Affairs*, vol. 10, pp. 185–246.

Dworkin, R. (1981b) "What is Equality? Part Two: Equality of Resources," *Philosophy and Public Affairs*, vol. 10, pp. 283–345.

Fleurbaey, Marc and Erik Schokkaert (2011) "Equity in Health and Health Care," *Handbook of Health Economics*, vol. 2, pp. 1003–92.

Fleurbaey, Marc and Vito Peragine (2013) "Ex Ante Versus Ex Post Equality of Opportunity," *Economica*, vol. 80, pp. 118–30.

Friedman, Milton (1953) "Choice, Chance and the Personal Distribution of Income," *Journal of Political Economy*, vol. 61, no. 4, pp. 277–90.

Friedman, Milton (1962) *Capitalism and Freedom*. Chicago: University of Chicago Press.

Friedman, Milton and L.J. Savage (1948) "Utility Analysis of Choices Involving Risk," *Journal of Political Economy*, vol. 56, no. 4, pp. 279–304.

Jusot, Florence, Sandy Tubeuf and Alain Trannoy (2013) "Circumstances and Efforts: How Important is Their Correlation for Measurement of Inequality of Opportunity in Health," *Health Economics*, vol. 22, iss. 12, pp. 1470–95.

Kanbur, Ravi (1987) "The Standard of Living: Uncertainty, Inequality and Opportunity," in G. Hawthorn (ed.) *The Standard of Living*, Cambridge: Cambridge University Press, pp. 59–69.

Kanbur, Ravi (2006) "The Policy Significance of Inequality Decompositions," *Journal of Economic Inequality*, vol. 4, no. 3, pp. 367–74.

Le Grand, J. (1987) "Equity, Health, and Health Care," *Social Justice Research*, vol. 1, no. 3, pp. 257–74.

Lim, Stephen S., Theo Vos, Abraham D. Flaxman, Goodarz Danaei, Kenji Shibuya, Heather Adair-Rohani, Mohammad A. AlMazroa et al. (2012) "A Comparative Risk Assessment of Burden of Disease and Injury Attributable to 67 Risk Factors and Risk Factor Clusters in 21 Regions, 1990–2010: A Systematic Analysis for the Global Burden of Disease Study 2010," *The Lancet*, vol. 380, no. 9859, pp. 2224–60.

Mirrlees, J.A. (1971) "An Exploration in the Theory of Optimum Income Taxation," *Review of Economic Studies*, vol. 38, pp. 175–208.

Ooghe, E., E. Schokkaert and D. Van de gaer (2007) "Equality of Opportunity Versus Equality of Opportunity Sets," *Social Choice and Welfare*, vol. 28, no. 2, pp. 209–30.
Peragine, Vito (2004) "Measuring and Implementing Equality of Opportunity for Income," *Social Choice and Welfare*, vol. 22, pp. 1–24.
Roemer, John (1998) *Equality of Opportunity*. Cambridge, MA: Harvard University Press.
Roemer, John, et al. (2003) "To What Extent Do Fiscal Systems Equalize Opportunities for Income Acquisition Among Citizens?," *Journal of Public Economics*, vol. 87, pp. 539–65.
Roemer, John (2008) "Equality of Opportunity," in S. Durlauf and L. Blume (eds), *The New Palgrave Dictionary of Economics*. Basingstoke: Palgrave Macmillan.
Swift, A. (2005) "Justice, Luck and the Family: Normative Aspects of the Intergenerational Transmission of Economic Status," in S. Bowles, H. Gintis, and M. Osborne-Groves (eds), *Unequal Chances: Family Background and Economic Success*. Princeton, NJ: Princeton University Press, pp. 256–76.
Wagstaff, Adam (2013, January 3) "Some Thoughts on Human Development, Equal Opportunity, and Universal Coverage." Blog post on Let's Talk Development. http://blogs.worldbank.org/developmenttalk/some-thoughts-on-human-development-equal-opportunity-and-universal-coverage.
World Bank (2005) *World Development Report 2006: Equity and Development*. Washington, DC: Oxford University Press for the World Bank.

Comments on "How Useful Is Inequality of Opportunity as a Policy Construct?" by Ravi Kanbur and Adam Wagstaff

Aristomene Varoudakis
University of Strasbourg, France

The paper by Kanbur and Wagstaff on the usefulness of inequality of opportunity in policy analysis is very relevant and balanced. It rightly acknowledges that the empirical analysis on the inequality of opportunity (IO) has contributed significantly to our understanding of the determinants of inequality of outcomes. It also correctly acknowledges the contribution of the IO concept to advancing the policy debate on income distribution, beyond what would have been possible by focusing only on the inequality of outcomes or on social equity and fairness concepts.

At the same time, the paper rightly emphasizes some of the concerns associated with the IO empirical analysis as developed by de Barros et al. (2009). First, the concept provides a lower bound estimate of the "true" inequality of opportunity, which is surrounded by considerable ambiguity. Second, the IO concept may delegitimize concern with the other components of inequality of outcomes – those related to differences in effort, luck, or talent.

Indeed, this last point might have been emphasized more: If there is inequality of opportunity, it may not be appropriate to assume that inequality of outcomes (due to differences in effort, luck, or talent) is fair and, therefore, legitimate from a policy perspective. The reason is that individuals that apply the same level of effort may get very unequal rates of return depending on the inequalities of opportunities they face.

This will be the case, for example, if capital or credit markets are imperfect, so that lack of collateral, reflecting inequality of opportunity (not differences in effort, luck, or talent), reduces individual borrowing ability. Such imperfections, owing to asymmetry of information, are well-known obstacles to the efficient functioning of credit markets (Stiglitz, 1989). Credit market imperfections may prevent these individuals, if they are farmers, for example, from upgrading the crops they grow to higher-quality crops, or from investing to expand their business if they are small entrepreneurs. Individual luck may also depend on inequalities of opportunity, if such inequalities prevent, for

example, individuals from improving educational attainment because they are stuck with second-class teachers in poor neighborhood schools. The effectiveness of individual efforts may then depend on inequality of opportunity to an extent surpassing what is captured by the empirical estimation of the IO measure.

The paper also makes the valid point that when outcomes depend on actions of third parties, the distinction between outcomes related to IO and to own actions become blurred. This is, for example, the case of individual health, given that immunization or preventive care depends on the motivation and efforts of parents.

An additional point perhaps worth emphasizing concerns the validity of international comparisons of the extent of inequality of outcomes attributable to the inequality of opportunity. To what extent, for example, does the higher contribution of IO to income inequality in Guatemala or in Colombia reflect more unequal circumstances or more imperfect capital and credit markets? More imperfect credit markets may rely more excessively on collateral, which, in turn, may be restricted for many individuals due to IO and initial circumstances beyond their control. What would be the main contributor to income inequality? Inequality of opportunity, or credit market imperfections, which may also differ substantially across countries? The diagnosis has important implications for the appropriate focus of policy to combat inequality.

Finally, the point made by the paper that the policy dialogue on inequality may lose traction if IO is quantified at a level that may seem of low significance to policymakers is perhaps exaggerated. There is research suggesting that growth is negatively affected by IO – as demonstrated by Marrero and Rodriguez (2010) for growth and IO in 23 US states. More generally, additional research seems to be needed on the interconnections, or loop, between inequality of opportunity, growth, and inequality of outcomes. The first part of the loop concerns the association between IO and growth: it is important to understand if IO has a robust negative incidence on growth by restricting the ability of otherwise equally skilled individuals to realize their productive potential. The second part of the loop is about the inclusiveness of growth: how growth and the institutions and policies that support it are associated with inequality of outcomes? The last part of the loop concerns the association between inequality of outcomes and IO: There is evidence, for instance from EU countries, that IO is positively associated with long-term unemployment and negatively associated with social protection expenditures (Marrero and Rodriguez, 2012). Understanding how the loop between IO, growth, and inequality of outcomes works would be essential for a sound, evidence-based dialogue on public policy, inequality, and growth.

References

de Barros, R., F.H.G. Ferreira, J.R. Molinas Vega and J. Saaverda (2009) *Measuring Inequality of Opportunity in Latin America and the Caribbean*. Washington, DC: World Bank.

Marrero, G.A. and J.G. Rodriguez (2010) "Inequality of Opportunity and Growth," ECINEQ Working Paper 2010–154, February. http://www.ecineq.org/milano/WP/ECINEQ2010-154.pdf.

Marrero, G.A. and J.G. Rodriguez (2012) "Inequality of Opportunity in Europe," *Review of Income and Wealth*, vol. 58, no. 4, pp. 597–621.

Stiglitz, J.E. (1989) "Financial Markets and Development," *Oxford Review of Economic Policy*, vol. 5, no. 4, pp. 55–68.

5

Toward a New Definition of Shared Prosperity: A Dynamic Perspective from Three Countries*

Hai-Anh H. Dang
World Bank

Peter F. Lanjouw
VU University, Amsterdam

5.1 Introduction

By the standards of a very austere international poverty line, such as the World Bank's $1.25 per person per day in 2005 Purchasing Power Parity (PPP) dollars, global poverty has fallen rapidly in recent decades (World Bank, 2015a). In many countries of the world, absolute poverty defined in these terms no longer affects significant segments of the population. This is a remarkable achievement that is rightly celebrated – even though it is clear that in certain countries and certain parts of the world, extreme poverty by this standard remains both widespread and stubbornly resistant to change. But poverty is not only thought of in absolute terms, and on the basis of an international standard. Most countries of the world assess poverty in their societies on the basis of national poverty lines that are largely anchored to the standards, expectations, and aspirations of their own societies. With social progress and economic growth, these standards typically evolve and consequently the poverty thresholds underpinning national poverty analysis also tend to rise (Ravallion and Chen, 2011). In this context, attention in many countries is shifting away from merely a focus on the rate of income growth among the poorer population groups towards also conducting an assessment of the quality of this growth. In particular, a key question that resonates in many countries is whether the poor are able to participate to the same degree and extent as the non-poor in a given country's growth process; whether they are sharing equally in the country's rising

*We would like to thank Sudhir Anand, Mario Negre, and participants at the Roundtable at the 17th World Congress of the International Economics Association (Jordan) for helpful discussions. We would like to note that Sudhir Anand's comments are based on an earlier draft presented at the Roundtable. We would also like to thank DFID for financial support through its Knowledge for Change and Strategic Research Programs.

prosperity. In an attempt to provide a quantifiable measure that engages with these concerns, the World Bank has recently proposed a definition of "shared prosperity" as growth in the income of the bottom 40 percent of the income distribution over time (for example, Basu, 2013; Jolliffe et al., 2015).[1]

This measure has many attractive features, most notably that it is easy to understand, can be straightforwardly estimated from household survey data, and has a historical pedigree within development thinking that dates back to the 1970s.[2] However, as with other welfare measures that attempt to collapse complex and multifaceted distributional outcome information into a single summary index, there are conceptual and measurement-related subtleties that the measure does not fully capture. For example, as was suggested above, there are good grounds to believe that at the country level, poverty remains an important concern for policy makers, even if the standards by which poverty is assessed are country-specific and are far from immutable over time as the country develops. It seems plausible that a debate about shared prosperity within a country would want to be able to refer to national-level poverty, even if it is understood that the notion of shared prosperity extends beyond a focus on poverty. Questions have been raised about the income threshold used in the World Bank's shared prosperity index to identify the population segment of interest: why should the income threshold be set at the 40th percentile, rather than say, the 20th percentile or 35th percentile? Why not instead use the percentile that derives from application of the country's own poverty threshold?[3] As discussed above, countries may vary in defining their national poverty lines, which in turn have a significant bearing on poverty rates. Why apply a blanket 40 percent to all countries? Furthermore, one might also ask why we do not consider growth of the income distribution as a whole instead for a more comprehensive analysis.[4]

Another limitation of the World Bank measure is that it focuses on the level of growth for the bottom 40 percent, rather than dynamic changes to the population shares of poor and vulnerable groups, since the target population is always fixed at this specified proportion. If there is an interest to track shared prosperity in terms of shifts in such population shares it would appear useful to develop a measure that focuses on such transitions explicitly.

We develop in this note an alternative approach to tracking changes in shared prosperity that is more closely anchored to traditional analysis of poverty. We employ a dynamic approach that considers not only the currently poor but also takes into account that segment of the population currently nonpoor but facing a heightened risk of falling back into poverty. We postulate that a process of growth that fails to reduce not only poverty but also that fraction of the population vulnerable to falling into poverty, cannot be regarded as representing fully satisfactory progress in boosting shared prosperity. Our proposed measure aims to circumvent the issues raised with respect to the

current measure of shared prosperity, discussed above, by clearly delineating the population into three income groups – poor, vulnerable, and secure (or "middle class") – and tracking how their population shares evolve. We examine the implication of alternative growth scenarios for the relative size of these groups over time. In this approach, shared prosperity is most obviously boosted when population shares of the poor and vulnerable are seen to decline, with a corresponding increase in the share of the population that can be viewed as secure. We present a simple, but perhaps practically useful, way of interpreting and ranking the different possible scenarios in terms of changes in shared prosperity.

We emphasize that our proposed measure is not intended to detract from, or supplant, the current approach of considering the growth of the bottom 40 percent. Rather, it is meant to offer an additional perspective that can help enrich our understanding. These two approaches are complementary and simply focus on different aspects of growth for population groups that fall in the lower part of the income distribution. As will be seen later, our approach is perhaps somewhat more complex and requires more detailed analysis, but in our examples it appears to generate results that are quite consistent with those deriving from the former approach.

To operationalize our analysis, we build on a method recently proposed by Dang and Lanjouw (2014) to construct vulnerability lines that, together with existing national poverty lines, can help identify the poor and the vulnerable in each country. These vulnerability lines are associated with given vulnerability indexes – defined as the percentage of the population that are currently non-poor but who face a significant risk of falling into poverty in the next period. For each country a socially acceptable vulnerability index has to be pre-specified – in the same way that countries have to specify national poverty lines that resonate within their respective societies.[5] Our method allows for shared prosperity to be defined both anonymously (i.e., for poor or vulnerable households in each period regardless whether they are different or the same) and non-anonymously (i.e., for the same poor or vulnerable households in the first period). Importantly, although the method focuses on dynamics and transitions in and out of specific population groups, it does not depend crucially on the availability of panel data. Instead, the method can build on methods to construct synthetic panels constructed from multiple rounds of cross-sectional data. The latter are far more frequently available in the developing country context and as a result, our method can potentially be applied to most developing countries.

For illustration, we analyze shared prosperity based on both the established approach, and our new proposed measure, using data from three countries from different income levels and geographical locations: India, the US, and Vietnam. Data from the US and Vietnam are actual panels respectively from the

Panel Study of Income Dynamics (PSID) and the Vietnam Household Living Standards Survey (VHLSSs), while data from India are synthetic panels that are constructed from cross-sections of the National Sample Surveys (NSSs). To maximize comparability across countries, we focus on the same time period for all three countries: 2004–08 for the US and Vietnam; and 2004/05–2009/10 for India.[6] Detailed calculations and estimation results are drawn from two companions to the present paper (Dang and Lanjouw, 2014, 2015).

This note consists of four sections. We briefly review the growth experience for India, the US, and Vietnam in section 5.2, before delving into the analysis of shared prosperity in section 5.3. Section 5.4 concludes with further thoughts on research directions.

5.2 Overview on growth experience of three countries

Table 5.1 provides two measures of economic growth, one is growth in GDP per capita, and the other is growth in household consumption per capita (total household income for the US). The latter measure is shown both for the whole population and the bottom 40 percent. While these measures can be qualitatively similar, they can offer different estimates of the speed of growth (for example, Deaton, 2005). Thus, combining the two can provide a more comprehensive picture. Indeed, Table 5.1 shows that growth rates are different, depending on whether the first or the second measure is employed.

A couple of observations are in order for this table. First, India had the largest annual growth rate for GDP per capita, which is then followed by Vietnam and the US for the considered periods. This order is different, however, when growth is defined in terms of household consumption. By this measure, overall growth rates in India and Vietnam were the same, and were higher than in the

Table 5.1 Growth experience of India, the United States, and Vietnam (percentage)

Annual growth rate	**Country**		
	India	**United States**	**Vietnam**
1. Based on national account			
GDP/ capita	6.6	1.3	5.8
2. Based on household survey			
Consumption/ capita for all the population	2.2	–1.8	2.2
Consumption/ capita for the bottom 40 percent	2.0	–1.3	3.1

Note: Annual growth rate is between 2004 and 2008 for the US and Vietnam, and 2004 and 2009 for India. GDP per capita data are from the World Development Indicators database. Survey-based consumption figures are from the NSS and VHLSS surveys respectively for India and Vietnam; for the US, this figure is total household income from the PSID. All survey-based numbers are estimated with population weights.

US. Second, while GDP per capita for the US had positive growth over this time period, survey-based income growth was even negative for the US. This suggests that this economy might in fact be contracting between 2004 and 2008, in stark contrast to the case in the other two countries. Finally, using growth of the bottom 40 percent as a measure of shared prosperity, Vietnam moved up and took the lead at 3.1 percent, followed by India at 2.0 percent, and finally the US at –1.3 percent. If comparison is made not across the countries, but within each respective country relative to the population as a whole, growth for the bottom 40 percent is stronger for Vietnam while it is weaker for India. On the hand, this growth is less negative than that of the whole population for the US (i.e., the mean consumption for the bottom 40 percent decreases less than that of the whole population).

We will come back to more discussion on each country in Section III.3. We discuss next the framework of analysis before considering a more disaggregated analysis of growth.

5.3 Analysis for shared prosperity

A key feature with our proposed approach to measuring shared prosperity is that we construct a vulnerability line, which separates out the non-poor population into two groups: those that are currently non-poor but face a heightened risk of falling into poverty in the next period, and those that are secure (and can be denoted middle class).[7] Since this vulnerability line is built upon the existing poverty line, employing these two lines together can provide a more comprehensive and more consistent analysis.

5.3.1 Vulnerability line: definition and estimation[8]

Let y_t and Z_t represent the household's consumption and the poverty line respectively in time t, t=0 and 1.[9] We define V_0 as the vulnerability line such that a specified proportion of the population with a consumption level *higher* than the poverty line but still *below* this line in time 0 will fall below the poverty line Z_1 in time 1. We designate the likelihood among this population of falling back into poverty in period 1 as the "vulnerability" index.[10]

We thus define the new vulnerability line as one that satisfies the following equality, given a specified vulnerability index $\mathcal{P}$

$$\mathcal{P}=P(y_1\leq Z_1|Z_0<y_0<V_0) \tag{1a}$$

or its equivalent expression,

$$\mathcal{P}=\frac{P(y_1\leq Z_1\cap Z_0<y_0<V_0)}{P(Z_0<y_0<V_0)} \tag{1b}$$

It can be useful to highlight some features of this vulnerability line. First, just as a poverty line can be constructed anchored to a benchmark (for example, level of energy or median household consumption), a vulnerability line can be constructed given a specific value for the vulnerability index $\mathcal{P}$ (say, 5 or 10 percent). Second, also similar to the poverty line, a lower value for this index is desirable and implies that a lower proportion of the population is at risk of falling into poverty.

However, a major difference between this vulnerability line and the poverty line is that the former is constructed using a dynamic poverty framework while the latter a static one; another is that this vulnerability line is defined to be used at the population level for population-averaged quantity rather than at the household level. Put differently, the construction of vulnerability lines is a two-step process. In the first step, (absolute) poverty lines are constructed, often based on minimum levels of calorie requirements. Then in the second step, these poverty lines provide a building block, which is then supplemented with information on the shares of the population defined in relationship to these poverty lines in both periods, to construct vulnerability lines.

Finally, this vulnerability line can offer a lower bound for the middle class. In other words, it can work as a lower bound value where households with a higher consumption than this line would be considered as belonging to the middle class, and households with a consumption level in between this line and the poverty line belonging to the group that is most vulnerable to poverty.

In terms of estimation, equality (1b) lends itself to straightforward estimation using household panel survey data, where the denominator can be estimated from the cross-section in time 0, and the numerator from the panel data spanning both time 0 and time 1. There is no closed-form solution for V_0 in equalities (1a) and (1b). However, given household consumption in both periods, the poverty line Z_1, and a pre-specified value for either the insecurity or vulnerability index, we can empirically solve for the vulnerability line V_0. In particular, since $\mathcal{P}$ is a decreasing function of V_0, we can iterate from the poverty line upward until we reach a value for V_0 that provides the specified vulnerability index. Given appropriate adjustments for inflation rates, the vulnerability line in time 0 can then be updated for later periods just as with poverty lines. In contexts where actual panel data are not available, synthetic panels can be constructed to substitute for such data (Dang, Lanjouw, Luoto, and McKenzie, 2014; Dang and Lanjouw, 2013).

For illustration purposes, Table 5.A1 in the Appendix shows the different vulnerability lines that correspond to different vulnerability indexes for India during the period 2004/05–2009/10. Further examples are provided in Dang and Lanjouw (2014, 2015).

5.3.2 Typology of poverty and vulnerability dynamics

The estimated vulnerability line can then be combined with the existing poverty line to classify the population into three welfare groups: Poor, Vulnerable, and Middle class. We propose a simple typology of different growth scenarios for the three welfare groups in Table 5.2. To obtain a ranking for the different growth scenarios, we adopt the following pro-poor criterion: Growth for the Poor category is most prioritized; and between the Vulnerable category and the Middle class, growth for the former has more priority. As a result, there are in total six possible growth scenarios depending on whether (the population share for) each of the three categories is expanding or shrinking.[11] The first three scenarios relate to the reduction of the Poor category, while the remaining three scenarios concern the expansion of this category. Thus by our pro-poor definition, these first three scenarios indicate positive growth, and the remaining scenarios suggest negative growth.

Table 5.2 indicates that the most positive pro-poor growth scenario is that both the Poor and Vulnerable categories are reduced while the Middle class category expands (Scenario 1). This is also the best general economic growth scenario, where everyone – regardless of whether they are rich, vulnerable or poor – is on average better off. The second-best growth scenario is one where

Table 5.2 Typology of welfare transition dynamics over two periods

Scenario	Pro-poor Growth	Welfare Category			Notes
		Poor	Vulnerable	Middle class	
1	Strongest/ Most positive	–	–	+	reduced poverty and vulnerability, and expanded middle class
2	More positive	–	+	+	reduced poverty, increased vulnerability, and expanded middle class
3	Positive	–	+	–	reduced poverty, increased vulnerability, and shrunk middle class
4	Negative	+	–	+	increased poverty, reduced vulnerability, and expanded middle class
5	More negative	+	–	–	increased poverty, reduced vulnerability, and shrunk middle class
6	Weakest/ Most negative	+	+	–	increased poverty, increased vulnerability, and shrunk middle class

Note: The signs (–) and (+) respectively stand for decrease and increase. Pro-poor growth is defined as the dynamics that are most beneficial to the different categories in this order: Poor, Vulnerable, and Middle class.

only the Poor category becomes smaller, while both the Vulnerable and Middle class swell (Scenario 2). The worst pro-poor scenario is where both the Poor and Vulnerable increase while the Middle class is reduced and everyone on average loses (Scenario 6), which is the opposite of Scenario 1. All the remaining scenarios can be similarly classified based on the changes in the sizes of these three categories.

Some findings emerge from this simple typology. First, pro-poor growth is strongest – or shared prosperity is largest – when both poverty and vulnerability are reduced. Otherwise, reduced poverty coupled with increased vulnerability (Scenarios 2 and 3) can potentially result in unstable poverty reduction. The reason is rather straightforward: assuming the increase in the Vulnerable category is mostly due to those households that just escaped poverty, without strong social protection programs, there is no guarantee that these households may not fall back into poverty in the next period. Consequently, for sustainable growth and more shared prosperity, more attention should be focused not only on reducing poverty but also on decreasing vulnerability, or aiding the vulnerable population category that are currently nonpoor but face a high risk of falling into poverty.

Second, the ranking provided in Table 5.2 can provide some rough guideline for a preferred pro-poor growth order for the different growth trajectories. If the objective is to achieve shared prosperity, this ranking suggests that growth for the poor and vulnerable – in this order – should be most prioritized. While the best pro-poor growth scenario is generally consistent with general economic growth (Scenario 1), this may not hold for other pro-poor growth scenarios where the Middle class can either expand or contract (Scenarios 2 and 3). Similarly, the whole economy may grow on average but poor households may even sink deeper into poverty if the Poor category swells (Scenarios 4 and 5). This priority should be well noted if shared prosperity is to be interpreted as more or at least equal growth for the poorer groups.

Finally, the typology provided in Table 5.2 is general enough to be employed with different definitions of vulnerability lines. Even though we derive these vulnerability lines based on the approach in Dang and Lanjouw (2014) for the analysis in this paper, these lines can also be obtained using other approaches. For example, one option is to derive the vulnerability line, also based on the probability of falling into poverty but using a regression-based framework (Ferreira et al., 2013; Lopez-Calva and Ortiz-Juarez, 2014); other options are to simply use some absolute cutoff thresholds such as between $2 and $10 PPP dollars (Banerjee and Duflo, 2008) or between the 40th and 80th percentiles of the income distribution (Alesina and Perotti, 1996).

5.3.3 Welfare analysis

We provide the empirical illustration in Table 3, where estimates are constructed based on the results from Dang and Lanjouw (2014).[12] We briefly

review the most relevant studies about pro-poor growth in each country before discussing estimation results.

5.3.3.1 India

India witnessed its GDP per capita increasing by almost half (47 percent) and poverty decreasing by 21 percent during the period 2004–09 (World Bank, 2015b). Not much, however, is known about pro-poor growth for India during this period, but recent studies (for example, Datt and Ravallion, 2011; Ravallion, 2011) suggest that economic growth has generally had a negative impact on poverty rates starting from the early 1990s. Our earlier estimates (Table 5.1) indicate that growth for the bottom 40 percent is slightly smaller than that of the whole population.

Estimation results shown in Table 5.3 confirm that growth in India in this period has been pro-poor, with the population share of the Poor category decreasing by 14 percent. However, this rate of decrease is slower than the growth rate of the Middle class at 19 percent. This period also saw the Vulnerable category expanding by 5 percent. The growth scenario for India in this period is definitely a positive case, and is second only to the most positive case of both reduced poverty and vulnerability and expanded middle class (Table 5.2).[13]

5.3.3.2 The United States

The year 2008 marks the Great Recession in this country, where subprime housing mortgages had a detrimental domino effect on other sectors of the economy. Petev, Pistaferri, and Eksten (2011) document that while real per capita consumption declined monotonically until the middle of 2009, the decline of real per capita disposable income was significantly smaller. This finding was

Table 5.3 Welfare transition dynamics of India, the United States, and Vietnam (percentage)

Welfare category	**Country**		
	India	**United States**	**Vietnam**
Poor	–14.4	12.4	–28.5
Vulnerable	4.8	9.4	–26.5
Middle class	19.0	–2.4	22.3
Pro-poor growth scenario	More positive	Most negative	Most positive

Note: Welfare transition is between 2004 and 2008 for the US and Vietnam, and 2004 and 2009 for India. Households are considered to be in the vulnerable category if their probability of falling from this status in the first period into poverty in the next period is at least 20 percent. Estimates are derived from Table 8 and Table 2.4 in Dang and Lanjouw (2014), where data for the US and Vietnam are true panels with the PSID and VHLSS, and data for India are synthetic panels constructed from the cross sections in the NSS. All survey-based numbers are estimated with population weights.

corroborated by De Nardi, French, and Benson (2012), who also found that the drop in income for poorer households was in fact lower than that of other households (thanks to means-tested transfer programs from the government).[14] Both the negative growth rate for mean income per capita, and the negative growth for the bottom 40 percent (shown in Table 5.1) – even though the latter is slightly larger – are consistent with these findings.

Our estimates (Table 5.3) suggests that growth in the US in this period has been least pro-poor, with the population share of the Poor category increasing by 12 percent. At the same time, the Vulnerable category expanded by 9 percent, while the Middle class contracted by 2 percent. The growth scenario for the US is in fact the most negative case for pro-poor growth according to the typology shown in Table 5.2.

5.3.3.3 Vietnam

Vietnam has been enjoying a steady GDP per capita growth rate of almost 6 percent in the period 2004–08 (Table 5.1), which followed a preceding decade of strong growth. Poverty has been declining rapidly in this country, and decreased by one quarter over this period – from 20 percent in 2004 to 15 percent in 2008. Economic growth in the previous decade (the 1990s) has been found to be strongly pro-poor (Glewwe and Dang, 2011).

Our estimates (Table 5.3) suggests that growth in Vietnam in this period has been solidly pro-poor, with the population share of the Poor category decreasing by 29 percent. At the same time, the Vulnerable category also decreased by 27 percent, while the Middle class expanded by 22 percent. The growth scenario for this country is in fact the most positive case for pro-poor growth according to our proposed typology.

We thus show that for the three countries during the period under study, Vietnam represents the most positive pro-poor growth scenario and India represents the next most positive, while the US reveals the least positive scenario.[15] It should be noted that while both the cited bottom 40 percent measure and our measure provide consistent results for these examples, this may not always hold for other countries.

5.4 Conclusion

We propose in this short paper an alternative measure of shared prosperity, which divides the population into three income groups and which is based on the changes of the population shares of each of these groups over time. We also offer a typology of different pro-poor growth scenarios based on this measure. Our proposed measure does not attempt to replace the measure of shared prosperity as growth of the bottom 40 percent recently proposed by the World Bank, but rather aims at examining different aspects of growth. Our

proposed measure would involve more intricate analysis, but the payoff is that it can provide richer analysis.

Furthermore, several strengths of our proposed measure merit attention. First, it is constructed using the existing poverty line and the resultant vulnerability line. As a result, it helps avoid the complicating issues of associating the bottom 40 percent with any existing national (or international) poverty line. Second, this measure emphasizes the importance of taking into account not only the poor but also the vulnerable (i.e., those that are currently non-poor but face a significant risk of falling into poverty) into the estimation of shared prosperity. Finally, the vulnerability lines used to construct this measure can be derived using a vulnerability index approach (Dang and Lanjouw, 2014) or other approaches. It is rather straightforward to estimate vulnerability for the former approach based on synthetic panels constructed from cross-sections.

It is useful to explore the combination of both approaches for analysis of shared prosperity, as illustrated in this paper. Estimation results using both approaches are qualitatively consistent for the three countries examined, and suggest that growth for Vietnam has the most shared prosperity, which is followed by India and the US. These results could well change for a more recent period, particularly since the period under study leads to the recent Great Recession period in the US.

We do not further investigate the dynamics of the movement between the different consumption categories for each country, which requires more detailed mobility analysis. But a promising direction for future research is to incorporate these between-group transitions into constructing a richer and more dynamic measure of shared prosperity. Another direction is to analyze the changes in shared prosperity not just for the poor as a whole, but also for specific subgroups among the poor and disadvantaged, such as the unemployed or those belonging to ethnic minorities.

Appendix

Table 5.A1 Vulnerability lines at given vulnerability indexes for India, 2004–2009

No	Vulnerability index (%)	Vulnerability line (rupee)	Increase (%)	Pop. share with consumption above poverty line but less than V-line (%)
1	35	508	5	3.4
2	34	528	9	6.2
3	33	543	12	8.2
4	32	553	14	9.4
5	31	578	20	12.6
6	30	598	24	15.0

(*continued*)

Table 5.A1 Continued

No	Vulnerability index (%)	Vulnerability line (rupee)	Increase (%)	Pop. share with consumption above poverty line but less than V-line (%)
7	29	623	29	17.9
8	28	648	34	20.7
9	27	673	39	23.3
10	26	703	46	26.2
11	25	743	54	29.8
12	24	783	62	33.1
13	23	823	70	36.1
14	22	868	80	39.1
15	21	923	91	42.3
16	20	998	107	45.6
17	19	1083	124	49.4
18	18	1213	151	53.4
19	17	1398	189	57.1
20	16	1723	257	60.6

Note: Vulnerability lines are in monthly rupees per capita in 2004 prices. The relative increases of the vulnerability line from the poverty line is shown under the column "Increase" (column 4). All numbers are estimated with synthetic panel data and weighted with population weights. The incremental value for iteration is 5 rupees. The exchange rate is US$1 for 45.3 rupees in 2004 (World Bank, 2015).

Table 5.A2 Welfare transition dynamics of India, the United States, and Vietnam, absolute changes (percentage)

Welfare category	Country		
	India	United States	Vietnam
Poor	–5.3	1.1	–5.7
Vulnerable	2.2	0.8	–6.6
Middle class	3.2	–2.0	12.3
Pro-poor growth scenario	More positive	Most negative	Most positive

Note: Welfare transition is between 2004 and 2008 for the US and Vietnam, and 2004 and 2009 for India. Households are considered to be in the vulnerable category if their probability of falling from this status in the first period into poverty in the next period is at least 20 percent. Estimates are derived from Table 8 and Table 2.4 in Dang and Lanjouw (2014), where data for the US and Vietnam are true panels with the PSID and VHLSS, and data for India are synthetic panels constructed from the cross sections in the NSS. All survey-based numbers are estimated with population weights.

Notes

1. In a slight abuse of notation, we use the terms income and consumption interchangeably in this note.
2. This measure was in fact proposed as early as 1972 in a speech by McNamara (1972), a former president of the World Bank, to the Board of Governors. A book subsequently published by the Bank (Chenery et al., 1974) provides more formal support. See Jolliffe et al. (2015) for more discussion on the historical development of this

measure; see also Currie-Alder et al. (2014) for a collection of papers on the evolution of development thinking.

3. It is perfectly possible, for example, if a country's national poverty rate is sufficiently below the cut-off point of 40 percent, for there to be an increase in poverty while growth in average income of the bottom 40 percent is judged to be acceptably high. Such an outcome could occur if inequality within the bottom 40 percent was increasing dramatically such that the poorest percentiles were seeing a decline in incomes while the group as a whole was seeing average incomes rise. It is unclear whether and how the concept of shared prosperity defined in this way would resonate within a country under these circumstances.
4. See Jolliffe et al. (2015, chapter 5) for further discussion of this point.
5. The vulnerability indices can be anchored to consideration of government budgetary constraints or to normative social welfare objectives. Also note that our proposed measure of shared prosperity is flexible enough to work with other definitions of the vulnerability line.
6. We use the 2007 and 2009 rounds of the PSID, whose income data are for the previous tax year. Thus we refer to these two rounds by the tax year for convenience.
7. The "middle class" defined in this way includes as well those households that would perhaps more reasonably be described as "rich". Note, however, that there is a general perception that household surveys do not typically capture well the richest segments of society. For example, Szekely and Hilgert (2007) show for the case of surveys from Latin America that the richest surveyed households generally reported incomes that were roughly similar or lower in dollar terms to what a middle manager in the US might earn. The authors argue that inequality calculated from households survey data in the LAC region is likely to be seriously underestimated as a result of these findings.
8. This section provides a brief overview of the vulnerability lines provided in Dang and Lanjouw (2014). Interested readers are encouraged to read the cited paper for more details.
9. We use the standard notation where y_t and Z_t are respectively a vector and a constant term; we also suppress the subscript for households to make notation less cluttered.
10. Dang and Lanjouw (2014) also offer another version of this vulnerability line which is based on an "insecurity index," where a specified proportion of the population with a consumption level *higher* than this line in time 0 will fall below the poverty line Z_1 in time 1. While sharing several common features with the vulnerability line provided here, this other version focuses on households in the top part of the consumption distribution.
11. Since these three groups add up to 100 percent, two other scenarios of either expanding or shrinking for all these groups as shares of the population are out of the question. In other words, the increases and decreases in the population shares of the three groups should cancel out each other in the total.
12. The figures shown in Table 5.3 are the relative changes of the different welfare categories. The absolute changes offer qualitatively similar results and are provided in Table 5.A2 in the Appendix.
13. Analysis of mobility for the more recent period (2009/10–2011/12) for India is provided in Dang and Lanjouw (2015).
14. In addition, Saez and Zucman (2014) document that over the 1986–2012 period, wealth per family averaged a growth rate of 1.9 percent per year but did not grow at all for the bottom 90 percent of US families.

15. However, note that if mobility between all income groups is considered, India has slightly more mobility than Vietnam, which is then followed by the US (Dang and Lanjouw, 2014).

References

Alesina, Alberto and Roberto Perotti (1996) "Income Distribution, Political Instability, and Investment," *European Economic Review*, vol. 40, no. 6, pp. 1203–28.

Banerjee, Abhijit and Esther Duflo (2008) "What Is Middle Class about the Middle Classes around the World?," *Journal of Economic Perspectives*, vol. 22, no. 2, pp. 3–28.

Basu, Kaushik (2013) "Shared Prosperity and the Mitigation of Poverty: In Practice and in Precept," Policy Research Working Paper No. 6700. Washington, DC: The World Bank.

Chenery, Hollis B., Montek S. Ahluwalia, C. L. G. Bell, John H. Duloy, and Richard Jolly (1974) *Redistribution with Growth: Policies to Improve Income Distribution in Developing Countries in the Context of Economic Growth*. London: Oxford University Press.

Currie-Alder, B., R. Kanbur, D. Malone, and R. Medhora (2014) *International Development: Ideas, Experience, and Prospects*. Oxford, UK: Oxford University Press.

Dang, Hai-Anh and Peter Lanjouw (2013) "Measuring Poverty Dynamics with Synthetic Panels Based on Cross Sections," World Bank Policy Research Working Paper No. 6504. Washington, DC: The World Bank.

Dang, Hai-Anh and Peter Lanjouw (2014) "Welfare Dynamics Measurement: Two Definitions of a Vulnerability Line and Their Empirical Application," World Bank Policy Research Paper No. 6944. Washington, DC: World Bank.

Dang, Hai-Anh and Peter Lanjouw (2015) "Poverty Dynamics in India Between 2004–2012: Insights from Longitudinal Analysis Using Synthetic Panel Data," World Bank Policy Research Paper No. 7270. Washington, DC: World Bank.

Dang, Hai-Anh, Peter Lanjouw, Jill Luoto, and David McKenzie (2014) "Using Repeated Cross-Sections to Explore Movements In and Out of Poverty," *Journal of Development Economics*, vol. 107, pp. 112–28.

Datt, Gaurav and Martin Ravallion (2011) "Has India's Economic Growth Become More Pro-Poor in the Wake of Economic Reforms?," *World Bank Economic Review*, vol. 25, no. 2, pp. 157–89.

Deaton, Angus (2005) "Measuring Poverty in a Growing World (Or Measuring Growth in a Poor World)," *Review of Economics and Statistics*, vol. 87, no. 1, pp. 1–19.

De Nardi, Mariacristina, Eric French, and David Benson (2012) "Consumption and the Great Recession," *Federal Reserve Bank of Chicago Economic Perspectives*, vol. 36, no. 1, pp. 1–16.

Ferreira, Francisco H. G., Julian Messina, Jamele Rigolini, Luis-Felipe López-Calva, Maria Anna Lugo, and Renos Vakis (2013) *Economic Mobility and the Rise of the Latin American Middle Class*. Washington, DC: World Bank.

Glewwe, Paul and Hai-Anh Dang (2011) "Was Vietnam's Economic Growth in the 1990's Pro-Poor? An Analysis of Panel Data from Vietnam," *Economic Development and Cultural Change*, vol. 59, no. 3, pp. 583–608.

Jolliffe, Dean, Peter Lanjouw, Shaohua Chen, Aart Kraay, Christian Meyer, Mario Negre, Espen Prydz, Renos Vakis, and Kyla Wethli (2015) *A Measured Approach to Ending Poverty and Boosting Shared Prosperity: Concepts, Data and the Twin Goals*. Washington, DC: the World Bank.

Lopez-Calva, Luis F. and Eduardo Ortiz-Juarez (2014) "A Vulnerability Approach to the Definition of the Middle Class," *Journal of Economic Inequality*, vol. 12, no. 1, pp. 23–47.

McNamara, Robert S. (1972) *Annual Address to the 1972 Annual Meetings of the Boards of Governors*. Washington, DC: World Bank.

Petev, Ivaylo D., Luigi Pistaferri, and Itay Saporta-Eksten (2011) "An Analysis of Trends, Perceptions, and Distributional Effects in Consumption," in David Grusky, Bruce Western, and Christopher Wimer (eds), *The Great Recession*. New York: Russell Sage Foundation, pp. 161–95.

Ravallion, Martin (2011) "A Comparative Perspective on Poverty Reduction in Brazil, China, and India," *World Bank Research Observer*, vol. 26, no. 1, pp. 71–104.

Ravallion, Martin and Shaohua Chen. (2011) "Weakly Relative Poverty," *Review of Economics and Statistics*, vol. 93, no. 4, pp. 1251–61.

Saez, Emmanuel and Gabriel Zucman (2014) "Wealth Inequality in the United States since 1913: Evidence from Capitalized Income Tax Data," NBER Working Paper No. 20625.

Szekely, Miguel and Marianne Hilgert (2007) "What's Behind the Inequality We Measure? An Investigation Using Latin American Data," *Oxford Development Studies*, vol. 35, no. 2, pp. 197–217.

World Bank (2015a) *Global Monitoring Report 2014/15*. Washington, DC: World Bank.

World Bank (2015b) *World Development Indicators Database*. Washington, DC: World Bank.

Comments on "Welfare Dynamics Measurement: Two Definitions of a Vulnerability Line and their Applications" by Hai-Anh H. Dang and Peter F. Lanjouw*

Sudhir Anand
Department of Economics, University of Oxford

This paper discusses a wide-ranging set of conceptual, statistical and empirical issues. It is concerned specifically with vulnerability to income poverty, and it identifies income cutoffs to define poverty and 'vulnerability' lines. These income cutoffs are then used to illustrate 'welfare dynamics' for panel surveys of households for the US and Vietnam, and for a *synthetic* panel constructed for India from separate rounds of cross-section data.

The authors' conceptual framework is based on a continuous bivariate distribution between the income of a person (household) today, Y_0, and the person's (household's) income tomorrow, Y_1. In keeping with standard statistical notation, I use upper-case letters for the random variables (Y_0, Y_1) and lower-case letters for realizations or values (y_0, y_1, z_0, z_1, v_0, v_0') of the random variables – in contrast to the notation in Dang and Lanjouw.

Their paper discusses vulnerability in terms of the probability of falling into poverty in period 1 *given* a person's (household's) income in period 0. The basic assumption made is that the higher one's income y_0 in period 0 the lower will be the probability of falling below any given threshold y_1 in period 1, and in particular the poverty threshold (or line) z_1 in period 1. This condition is known as "positive regression dependence" (Lehmann 1966, p. 1143, condition (5.3)). Positive regression dependence (PRD) of the random variable Y_1 on Y_0 is expressed formally as:

$$P(Y_1 \leq y_1 | Y_0 = y_0) \text{ is decreasing (non-increasing) in } y_0, \text{ for all } y_1.$$

This is Assumption 1 in Dang and Lanjouw, who call the condition "Y_1 is stochastically increasing in Y_0".

*These comments are based on the paper presented by Dang and Lanjouw with the above title at the IEA roundtable in Jordan.

Weaker and stronger conditions of positive dependence or association between the random variables Y_0 and Y_1 are also examined in Lehmann (1966).[1] They are all meant to characterize in different ways the notion that large values of Y_0 tend to be associated with large values of Y_1, and small values of Y_0 tend to accompany small values of Y_1.

In fact, Lehmann (1966) introduced three successively stronger conditions of positive dependence or positive association between the random variables Y_1 and Y_0, which, in order of increasing strength, are as follows:

(1) Condition (5.1) in Lehmann (1966, p. 1143), or positive quadrant dependence (PQD):

$$P(Y_1 \leq y_1 | Y_0 \leq y_0) \geq P(Y_1 \leq y_1) \text{ for all } y_0, y_1.$$

This condition expresses the fact that knowledge of Y_0 being small (less than y_0) increases the probability of Y_1 being small (less than y_1).

(2) Condition (5.2) in Lehmann (1966, p. 1143):

$$P(Y_1 \leq y_1 | Y_0 \leq y_0) \text{ is decreasing (non-increasing) in } y_0, \text{ for all } y_1.$$

This condition has been called left-tail decreasing (LTD) by Balakrishnan and Lai (2009, p. 111), although it is slightly mis-defined there. Balakrishnan and Lai (2009) also discuss the obverse condition called right-tail increasing (RTI), which is defined formally as:

$$P(Y_1 > y_1 | Y_0 > y_0) \text{ is increasing (non-decreasing) in } y_0, \text{ for all } y_1.$$

(3) Condition (5.3) in Lehmann (1966, p. 1143) of positive regression dependence (PRD) mentioned earlier:

$$P(Y_1 \leq y_1 | Y_0 = y_0) \text{ is decreasing (non-increasing) in } y_0, \text{ for all } y_1.^2$$

Lehmann (1966, Lemma 4, p. 1144) demonstrated that:

$$\text{PRD} \rightarrow \text{LTD} \rightarrow \text{PQD}.$$

It is also straightforward to show that:

$$\text{PRD} \rightarrow \text{RTI} \rightarrow \text{PQD}.$$

The assumption of PRD (or "Y_1 stochastically increasing in Y_0") invoked by Dang and Lanjouw (Assumption 1) implies the weaker conditions of LTD and RTI, which in turn imply PQD. The main propositions of their paper (Propositions 1–3) follow immediately from the conditional probabilities that

define LTD and RTI – by substitution of the poverty line z_0 (or the vulnerability lines v_0 or v_0' above z_0) for y_0, and the poverty line z_1 for y_1. Indeed, there is not much daylight between their Assumption 1 and the propositions stated and proved in their paper. The results are straightforward in view of the assumption made that the higher a person's (household's) current income y_0, the lower the probability that the person (household) will fall below any given threshold y_1 in period 1, e.g. the poverty line z_1.

Corresponding to a vulnerability line v_0 *above* z_0, the authors define an "insecurity index" $\mathcal{P}^1$ as $\mathcal{P}^1 = P(Y_1 \leq z_1 | Y_0 > v_0)$, and a "vulnerability index" $\mathcal{P}^2$ as $\mathcal{P}^2 = P(Y_1 \leq z_1 | z_0 < Y_0 \leq v_0)$. They show that under Assumption 1, $\mathcal{P}^2 > P(Y_1 \leq z_1 | Y_0 > z_0) > \mathcal{P}^1$ (Proposition 3).

Indeed, in their characterization of vulnerability solely in terms of a person's (household's) current income level – and *no* other variable – their Assumption 1 will also imply that the income group most vulnerable to poverty in period 1 is those who are *poorest* in terms of income (depth of poverty) in period 0. In particular, it follows that the *currently poor* are more vulnerable to poverty in period 1 than those defined by the authors as "secure" [sic] in their insecurity index $\mathcal{P}^1$ or "vulnerable" in their vulnerability index $\mathcal{P}^2$. In the spirit of Dang and Lanjouw, I offer Proposition A with its proof in Appendix A.

Proposition A:

Given LTD (condition (5.2) of Lehmann 1966), it follows that:

$$P(Y_1 \leq z_1 | Y_0 \leq z_0) > P(Y_1 \leq z_1 | z_0 < Y_0 \leq v_0) = \mathcal{P}^2$$

Hence in comparison to the groups identified by Dang and Lanjouw above the poverty line z_0 (viz. those between z_0 and v_0, those above z_0, and those above v_0), the most vulnerable group is those who are *currently poor*. In this framework, a concern with poverty and a distinct concern with vulnerability will lead to focussing on exactly the same group, viz. those who are currently income poor.

The question naturally arises, therefore, whether low income is all there is to vulnerability. Elsewhere in commenting on human security, I discussed vulnerability in terms of the magnitude of downside risk a person faces in relation to how much above the minimum threshold he or she is in the relevant dimension (Anand 2012, p. 9). I also noted that: "The extreme case of insecurity is certainty of being below a specified threshold, and the absence of any chance of avoiding that fate."

In a special contribution to the *Human Development Report 2014*, Stiglitz has argued that:

> If we are to formulate policies to reduce vulnerability, it is essential to take a broad view about what creates such vulnerability. Individuals and societies

> are inevitably exposed to what economists call 'shocks', adverse events that have the potential to lead to marked decreases in living standards. The larger the shocks, the greater their depth and duration, and the greater vulnerability, other things equal. … Vulnerability has multiple causes and consequences. … [I]f we are to succeed in reducing vulnerability, we need to approach it from a broad systemic perspective. (UNDP 2014, pp. 84–5)

In their empirical work, Dang and Lanjouw examine vulnerability (and its causes) only through *lines* in income space, which they then use for the purpose of studying income ('welfare') dynamics or mobility. Given the observed bivariate distribution of (Y_0, Y_1), the vulnerability line v_0 is drawn above the poverty line z_0 in period 0 such that the risk of falling into poverty in period 1, that is, the vulnerability index $\mathcal{P}^2 = P(Y_1 \leq z_1 | z_0 < Y_0 \leq v_0)$, is 0.1 or 10 percent. This vulnerability line is used to calibrate income intervals for examining mobility between three income groups in a population – the poor (those below z_0), the vulnerable (those between z_0 and v_0), and the 'middle class' (defined as those with income above v_0). The authors go on to estimate income dynamics among these groups in three countries: USA, Vietnam, and India.

But to study mobility one could equally define the income groups in terms of multiples of the poverty line in a country. Alternatively, one could, for example, define the income cutoffs through *quartiles* of the income distribution. However, if the object were to study vulnerability per se, and to identify groups that are vulnerable and the causes of their vulnerability, the exercise might be quite different from an identification of groups solely through particular income intervals.

To analyze vulnerability, one would need to identify groups more broadly through the downside risk they face which might push them into poverty. Particular occupational, socioeconomic and geographical groups may be more prone to downside risk than others. Examples might include farmers who are exposed to weather risk, people who are in ill-health and have no health insurance, people who live in locations subject to drought, floods and dangerous epidemiological environments, the disabled who are at greater risk of suffering from violence, disaster and other hardships, and so on.

Dang and Lanjouw also claim that their approach offers an "appealing basis for defining and identifying the middle class in society" – whom they define as everyone with income above the vulnerability line v_0. Their 'middle class' therefore encompasses the rich and the super-rich – including the top 1.0 percent. But again, if the purpose is to identify the middle class, sociologists and other economists would define this group in society differently – in terms of their level of education, or type of occupation, or sources of income, or middle position in the income distribution, or patterns of consumption (e.g. for advertising and marketing purposes), and so on. It is arguable whether the middle class should be defined in terms of an income group facing a particular probability

of falling below the poverty line. According to this criterion, if incomes were *stable* over time then *all* non-poor people would constitute the middle class – including those just above the poverty line as well as the rich. It is not clear what purpose is served by defining the middle class in this manner. The authors' definition of the middle class would seem to require more justification than they offer.

In conclusion, I confess to being a little unclear about the purpose of the definitions and statistical structures presented in this paper. If the motivation is to study income dynamics, the vulnerability line appears to have no more salience than specifying income cutoffs in other ways – e.g. as multiples of the poverty line, or as quartiles (or other percentile groups) of the income distribution. If it is to identify vulnerable groups, one would need to look at groups whose income is subject to substantial downside risk – which may not be a function solely of their current income level. If it is to identify the middle class, there are again likely to be more salient approaches to defining such a group. In my opinion, the paper would benefit from a sharper focus and a clearer justification for the analyses it presents.

Appendix A

Proposition A:

Given LTD (condition (5.2) of Lehmann 1966), it follows that:

$$P(Y_1 \le z_1 | Y_0 \le z_0) > P(Y_1 \le z_1 | z_0 < Y_0 \le v_0)$$

Proof:

Since $v_0 > z_0$, LTD implies

$$P(Y_1 \le z_1 | Y_0 \le z_0) > P(Y_1 \le z_1 | Y_0 \le v_0) = \frac{P(Y_1 \le z_1 \cap Y_0 \le v_0)}{P(Y_0 \le v_0)}$$

Now

$$P(Y_1 \le z_1 \cap Y_0 \le v_0) = P(Y_1 \le z_1 \cap Y_0 \le z_0) + P(Y_1 \le z_1 \cap z_0 < Y_0 \le v_0)$$

and

$$P(Y_0 \le v_0) = P(Y_0 \le z_0) + P(z_0 < Y_0 \le v_0)$$

Therefore,

$$P(Y_1 \le z_1 | Y_0 \le v_0) = \frac{P(Y_1 \le z_1 \cap Y_0 \le z_0) + P(Y_1 \le z_1 \cap z_0 < Y_0 \le v_0)}{P(Y_0 \le z_0) + P(z_0 < Y_0 \le v_0)}$$

$$= w_1 \, P(Y_1 \le z_1 | Y_0 \le z_0) + w_2 \, P(Y_1 \le z_1 | Y_0 \le v_0)$$

where

$$w_1 = \frac{P(Y_0 \leq z_0)}{P(Y_0 \leq z_0) + P(z_0 < Y_0 \leq v_0)}$$

$$w_2 = \frac{P(z_0 < Y_0 \leq v_0)}{P(Y_0 \leq z_0) + P(z_0 < Y_0 \leq v_0)}$$

and $w_1 + w_2 = 1$.

Therefore,

$$P(Y_1 \leq z_1 | Y_0 \leq z_0) > w_1 \, P(Y_1 \leq z_1 | Y_0 \leq z_0) + w_2 \, P(Y_1 \leq z_1 | z_0 < Y_0 \leq v_0)$$

so that

$$(1 - w_1) \, P(Y_1 \leq z_1 | Y_0 \leq z_0) > w_2 \, P(Y_1 \leq z_1 | z_0 < Y_0 \leq v_0)$$

But $(1 - w_1) = w_2$, and hence

$$P(Y_1 \leq z_1 | Y_0 \leq z_0) > P(Y_1 \leq z_1 | z_0 < Y_0 \leq v_0) \qquad \square$$

Notes

1. See also Balakrishnan and Lai (2009, chapter 3).
2. There is an even stronger condition of positive dependence than PRD, called "positive likelihood ratio dependence" (Lehmann 1966, pp. 1150–1), which requires that the conditional density of Y_1 given $Y_0 = y_0$ has a monotone likelihood ratio.

References

Anand, S. (2012): "Human Security and Universal Health Insurance," *The Lancet*, vol. 379, January 7, pp. 9–10.

Balakrishnan, N. and C. D. Lai (2009): *Continuous Bivariate Distributions*. New York: Springer.

Lehmann, E. L. (1966): "Some Concepts of Dependence," *The Annals of Mathematical Statistics*, vol. 37, no. 5, pp. 1137–53.

UNDP (2014): *Human Development Report 2014 – Sustaining Human Progress: Reducing Vulnerabilities and Building Resilience*. New York: United Nations Development Programme.

6

Behavioral Economics and Social Exclusion: Can Interventions Overcome Prejudice?*

Karla Hoff
World Bank

> Human beings can live, and human liberty can exist, only within a system of norms, meanings, and roles; but in any particular form, these things can impose severe restrictions on wellbeing and autonomy.
>
> Sunstein (1996, 911)

In many societies, broad segments of the population are barred from full human rights and full participation in society: they are "socially excluded" (World Bank 2013). Social scientists seek to understand the mechanisms by which exclusion occurs in order to determine how to mitigate it. The rational actor model assumes that individuals see all their options objectively, reason without bias, know what is in their self-interest, and act accordingly. The model implies that the dismantling of exploitative structures and unjust, formal barriers to certain groups in markets, schools, and neighborhoods are "all" that is required to end social exclusion.

However, behavioral economics has found that people are not always rational. Institutions have a "schematizing power" that affects information-processing and behavior (Bruner 1990, 58; DiMaggio 1997, 271). Mental models (or equivalently, schemas) that people have absorbed from institutions can distort perceptions in ways that reproduce social exclusion long after the unjust, formal barriers have been removed. "Spoiled collective identities" (a term due to Loury 2002) need to be repaired to overcome social exclusion.

Consider, for instance, discrimination against African-Americans. The rational phenomenon of "statistical discrimination" occurs when accurate, group-level estimates of difficult-to-observe characteristics – such as productivity, loyalty, or leadership qualities – provide information in assessing

*I am grateful to Allison Demeritt, Paola Giuliano, Priyanka Pandey, and James Walsh for helpful comments.

an individual's traits. An irrational phenomenon occurs when despite the facts, an individual is considered unworthy. A legacy of American slavery is a modern-day "spoiled collective identity" for African-Americans that perpetuates systematic disadvantage (Loury 2002, 60–1). When résumés with equal qualifications are submitted in job applications, candidates with names associated with African-Americans are less likely than candidates with names associated with whites to be viewed as qualified (Bertrand and Mullainathan 2003). When actual people present themselves to apply for low-wage jobs, African-American applicants *without criminal records* are offered jobs at a rate not significantly higher than the hiring rate for white applicants with identical résumés but *with criminal records* (Pager, Western, and Bonkowski 2009). Jail sentences for vehicular manslaughter are shorter when an African-American is killed than when a white is killed. Since the race of the victim may be random, the shorter jail sentences can be interpreted as a racial bias (Glaeser, Sacerdote, and Scheinkman 1996). When Prohibition produced high incarceration rates for whites, many people viewed the law as harmful, leading the United States to repeal the ban on alcohol (Kyvig 1979). But when U.S. laws against drug use produce high incarceration rates for African-Americans, many people view the outcomes as consistent with the stigmatized identity of African-Americans. Only recently have there been major efforts to change the US federal laws (see *Los Angeles Times,* August 11, 2013).

Human mental processing power – perception, attention, reasoning, and retrieval from memory – is much more limited than the rational actor model assumes. Humans rely on automatic thinking and mental shortcuts for much of their decision making. Potentially thousands of details could be observed at any given moment. Mental models are cognitive representations, often shared among members of a society, that affect where people direct their attention and how they structure information (see, for example, Axelrod 1973). The models provide default assumptions, means of categorizing perceptions, causal narratives, and associations that help individuals make sense of the world. There is a strong link between automatic thinking and perception (Kahneman 2003, 2011; Gennaioli and Shleifer 2010); both involve the construction of meaning in a process that the perceiver is generally unaware of. He imagines that he is responding objectively to the situation.

The transmission of mental models between persons and across generations gives the models life beyond the circumstances that originally gave rise to them. This can explain why past structures of power can have persistent effects on mental models and, thus, on perception.

Loury (2002, 37) asks a key question: If a representation attributed to a group, such as a race, gender, caste, or class, is a human product – that is, a social construction – then shouldn't humans be able to control it? Can humans intervene to give people new mental models or to alter which mental models

are activated in a particular context, in such a way that individuals judge each other more fairly and are able to respond more fully to their opportunities? This paper considers three interventions that have had some success in developing countries. It elaborates on themes of behavioral development economics in the *World Development Report 2015* (World Bank Group 2015). But first a brief primer on cognition is in order.

A primer on cognition

Psychologists, neuroscientists, and economists have been converging on a new understanding of the brain's functioning.[1] Humans use a dual process to make decisions – fast, intuitive thinking and slow, controlled thinking. Psychologists metaphorically label the two modes of thinking as distinct systems or distinct actors in the mind: System 1 is the *automatic system,* and System 2 is the *deliberative system* (Kahneman 2003, 2011). In most situations, people use automatic thinking unless provoked to check their thoughts deliberatively.

Automatic thinking makes implicit use of mental models (schemas) to process information. The use of a mental model may create neuronal connections that are not easily undone (Kitayama and Park 2010). When neurons are consistently activated by co-occurring features of experience, the connections among the neurons are strengthened:

> Thereafter, if one of those neurons is activated, it will be more likely to activate another in that group… These neural changes determine the pathways through which activation spreads until a particular response is evoked. The synaptic changes that make this happen cannot be erased like sentences from a text...
>
> [O]nce a network of strongly interconnected units has been created, it fills in ambiguous and missing information by activating all the units in an interconnected network, even those not directly stimulated by current experience..., reinforcing our original expectations… (Strauss and Quinn 1997, 90)

When the mental models that individuals use are well-adapted to the task at hand, they make individuals better off: "Time and energy are saved, rumination and doubt are reduced, and nothing important is lost" (Ross and Nisbett 1991, 77; see also Todd and Gigerenzer 2000). But in other cases, mental models lead people to make systematic mistakes. Automatic cognition is not optimizing:

> what our attention is drawn to, what we focus on, what we recall is not always what is most necessary or needed for optimal decision making. Some

> critical information is ignored; other – less relevant – information receives undue attention because it stands out… By governing what we are thinking about, [automatic thinking] shapes what we conclude, even when we are thinking hard (Shleifer 2012, 10).

Automatic thinking may be at the root of the poor outcomes of particular social groups. It could help explain, for example, patterns of racial and gender and class discrimination in the job market if what stands out to prospective employers are not a job-seeker's qualifications, but his stigmatized social identity.

Even seemingly insignificant aspects of context can affect judgments by focusing attention and triggering associations. In a test of implicit attitudes, individuals were asked whether they preferred Group 1 or Group 2, where Group 1 were well-liked, African-American athletes and Group 2 were disliked, white politicians (Mitchell, Nosek, and Banaji 2003). Respondents liked Group 1 more than Group 2 when the context emphasized members' *occupation*, but liked Group 2 more than Group 1 when the context emphasized members' *race*. In other words, the subjects reversed their preferences when the context primed a different social identity. Category names influenced the valence of the members of the categories. The findings suggest that automatic attitudes are contextually driven and malleable.[2]

As DiMaggio emphasizes, cultural understandings are fragmented and can be inconsistent. A contextual change that shifts which mental model is activated can change how individuals behave and thus the environment they collectively create. In DiMaggio's words, "large-scale cultural changes may be caused by large-scale, more or less simultaneous frame switches by many interdependent actors" (280). The environment and the mental models on which people draw are thus jointly determined.[3] Standard economics, based on the rational actor model, overlooks this close interrelationship between context and behavior and does not take into account the role of schematized thinking (i.e., thinking shaped by mental models) in the process of social change.[4] In contrast, sociologists emphasize that widely shared behaviors have a "schematizing power" and that "the psychology of mental structures provides a microfoundation to the sociology of institutions" (Bruner 1990, 58): *Society and its institutions rest on cognitive foundations*.

This paper discusses three kinds of interventions that have had some success in developing countries in changing the mental models that individuals used to process information and thereby reduced social exclusion. In the first two interventions, success entailed creating new mental models: new models (a) of female genital cutting and (b) of women political leaders. In the third intervention, success entailed making a stigmatized social identity less salient. Figure 6.1 represents the argument in a simple way.

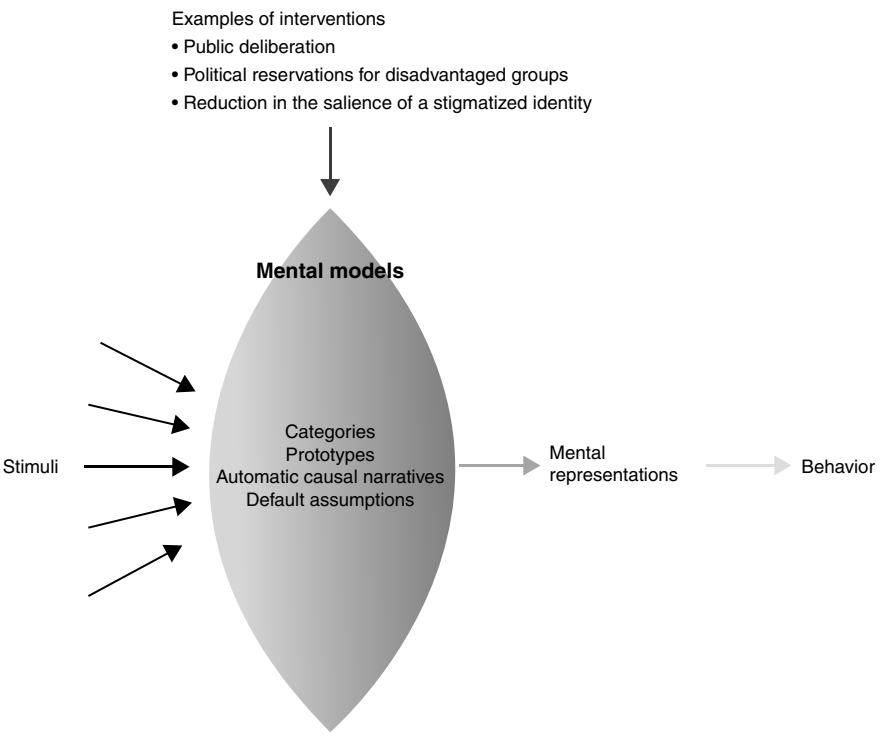


Figure 6.1 Interventions to change mental models that influence behavior
Note: People automatically use mental models to structure and interpret their experiences and guide their behavior. Interventions can create new mental models or change which model is activated in a given context. This paper focuses on the three types of interventions, shown in the figure.
Source: Author.

Intervention 1. Group deliberation

Prior beliefs can create blinders in reasoning. Group deliberation has the capacity to improve reasoning. Group deliberation is particularly important for social constructions, since no individual can change them on his own. For example, is a mother who educates her daughter a bad or foolish parent? Is it natural for young girls to bleed extensively or can such bleeding occur as a result of female genital cutting (FGC)? Are women leaders generally low quality?

This section begins by summarizing the evidence that beliefs can create biases in reasoning. Then it considers biased reasoning related to FGC and shows that group deliberation in Senegal reduced the biases and laid the foundation for the abandonment of the practice.

In cognitive science, reasoning means an inference at the conceptual level in which a new mental representation (a *conclusion*) is consciously produced

and the previously held representations (*premises*) that warrant it are also consciously entertained. An important limitation to reasoning is *confirmation bias,* the tendency to search for, interpret, and remember information in a way that supports one's initial beliefs.[5] Mercier and Sperber (2011) argue that the primary function of reasoning is to win arguments, not to improve knowledge: when individuals have some initial intuition or hypothesis, reasoning is used not to evaluate the initial intuition but to find justifications for it. An individual's beliefs shape what the individual pays attention to, what he remembers, and how he interprets a situation. An individual's beliefs can shape what information he "rationalizes away," so that initial beliefs can be difficult to dislodge even when the data would seem, objectively, to refute them. People often mistakenly believe that their use of reasoning is helping them make better decisions when, in fact, it is being guided by an irrational loyalty to initial beliefs and a tendency to discount, misinterpret, forget, or avoid information that does not support them. The irrational loyalty to one's beliefs is an asset in an argumentative context because it makes the individual good at discovering weaknesses in the arguments of others. As long as the deliberators seek the truth and there is some diversity of beliefs among them, the group can engage in a highly efficient division of cognitive labor. If one person proposes view A and another person counters with view B, each will specialize in the pros of his own view and the cons of the other's view.

Experiments reviewed by Mercier and Sperber show that deliberation among people who want to learn the truth leads to more efficient outcomes than individual thinking does: "Reasoning should produce its best results when used in argumentative contexts, most notably in group discussions" (Mercier and Sperber, 61). Reasoning in groups, as long as there are differences in the intuitions of the members of a group, can serve epistemic goals. One class of experiments compares performance in group versus individual settings in tasks for which there exists a demonstrably correct answer. In the most widely used task, the percentage of correct answers under individual performance was 10 percent but rose to 80 percent under group performance. In the groups, truth wins; as soon as one participant has understood the problem, he can convince the group that his solution is correct. Often the performance of the group is better than that of its best member, as several participants may be partly wrong and partly right. Even large monetary incentives do not improve performance in these tasks. It is the group setting that elicits a high level of performance in the outcome of reasoning.

Group deliberation can provide a forum in which dissenters from a prevailing way of thinking can discuss beliefs about nature that tend to be taken for granted and therefore that many people have not examined critically. Group deliberation can also raise awareness of inconsistencies in symbolic systems that individuals, reasoning on their own, would be unlikely to notice. But

group deliberation is not a panacea; it could also result in heightened commitment to a particular way of thinking and make "deviant" symbolic systems dangerous for an individual to possess. The next section describes a success case in which group thinking changed reasoning that had justified the widespread practice in Senegal of female genital cutting.

Case study: The abandonment of female genital cutting in Senegal

Beliefs play a much larger role in behavioral economics than in standard economics. In behavioral economics, beliefs shape what is perceived, which means that there is much more for groups to deliberate about than merely their own or others' self-interest. This section discusses the abandonment in Senegal of FGC. Group deliberation changed perceptions and overcame biases in reasoning and so changed the social meanings that supported this practice. While FGC had been practiced for centuries, it was abandoned in a single generation in many parts of Senegal.

Mackie (2003) reports several features of FGC in Senegal:

> It is very widely practiced in many communities.[6]
> It is supported and transmitted by women across generations.
> It is believed to ensure female chastity and fidelity.
> It is believed to promote women's health and fertility.
> It can promote proper marriage and family honor and enhance the status of the woman.

Biased perceptions and reasoning help to maintain the practice. Writing about complications of infibulation in Islamic northeast Africa, Hicks (1993, 73) observed:

> Women do not even correlate subsequent physical discomfort, pain, and related gynecological and obstetric problems with having been circumcised. Such physical problems are perceived as being the common lot of women. This is because the problems are, to one degree or other, prevalent among the majority of infibulated women; they are not viewed as unusual. Logically then, neither the act of infibulation nor related sequelae (unless requiring emergency treatment) are high priority issues for women in these societies.

Writing about events in a Fulani village in Senegal, Mackie (2003, 147–8) noted:

> On hearing of the causal relationship from a source they considered credible… it took (a group of local women) thirty minutes of discussion to decide that the causal claim was correct. They reviewed local history and suddenly

> realized that incidents of death, haemorrhaging, and infection were immediately associated with [FGC], and they broke down and wept. One woman told me that she had her daughter who had haemorrhaged seriously stand next to a girl of the same age who was taller by about a foot. "She's never been the same since the cutting," I was told. "Before she ran around all day and played and since she's been quiet and dull."

Parents want to protect their children. At the same time, a broadly shared ideal among many communities in Senegal was that parents should have their daughters cut, and that parents should marry their sons only to girls who had been cut. The non-governmental organization Tostan organized discussions within small, fixed groups for periods of two to three years (Cislaghi, Gillespie, and Mackie 2014). Generally, discussions within a fixed group of individuals occurred two or three times a week. A trained facilitator employed by Tostan led discussions in each group about human rights. For individuals in this area, a period of reflection was needed to understand FGC as a violation of rights to life, health, and bodily integrity. Mackie reports in his field notes a discussion from one group meeting:

> A Bambara group was told the story of Chinese footbinding by their non-formal education facilitator. The participants thought it was horrifying that parents would do such a thing to their children. The facilitator responded that Europeans looked on the parents who do FGM/C [female genital mutilation/cutting] in the same way. "No, no, no," the participants responded. "We do this to help our daughters." "So did the Chinese," the facilitator said. (Mackie field notes, quoted in Mackie and Lejeune 2009, 21)

Group deliberation led to a process of questioning by fostering a change in perceptions. It led to the recognition of the inconsistencies between one belief (that one should foster children's wellbeing) and another (that girls should be genitally cut). Once the recognition occurred, a coordination problem emerged in intermarrying communities. To solve it required organizing commitments for the collective abandonment of the practice. There were thus two critical stages to abandon a strong norm of FGC: (1) group deliberation on perceptions and beliefs; and (2) coordination of actions within the intermarrying community.

Intervention 2. Political affirmative action

In the rational actor model, perceptions are objective and autonomous; each thinker is a sovereign individual. The theory cannot account for the collective, classificatory frameworks within which individuals choose. The anthropologist

Mary Douglas (1986) argues that institutions teach people how to "see," how to "think," what sets of things are "similar," and what are the important categories. Institutions operate not only as patterns of activity and "rules of the game," but also as symbolic systems (Friedland and Alford 1991). A society may have inconsistent symbolic systems, but individuals may not be aware of the inconsistencies (Swidler 1986). A policy that temporarily mandates inclusion of a group might change the commonly prevailing mental model and thereby reduce social exclusion. This section presents two case studies of such policies. The first was successful and the second was not, at least not in the short run.

Case study of a success: The effect of political affirmative action for women in West Bengal, India

A 1993 amendment to the constitution of India made it mandatory for state governments to reserve for women in one-third of the villages the position of village head (*Pradhan*). The Indian state of West Bengal began implementing the amendment in 1998. West Bengal randomly assigned the reservation across villages.

What were the results? There is no evidence that reservations for women lowered the quality of governance, but much evidence that exposure to women *Pradhans* changed mental models in ways that reduced the social exclusion of women (Beaman et al. 2009, 2012). The experience of living under a woman *Pradhan* erased the prejudice, on average, of male villagers against women leaders by several measures – an Implicit Association Test, the evaluation of political speeches, and the assessment of the quality of actual village *Pradhans*. The experience of living under a female *Pradhan* reduced the gap between parents' aspirations for sons and daughters. In villages that had women leaders, parents' expectations for their daughters were higher, and girls have gone to school longer and had slightly fewer hours of housework. After the reservations ended in a village, women have run for political office in higher percentages and in many cases won the elections, as shown in Figure 6.2.

The greater presence of female political representatives produced a surprising change in women's reporting of crimes against women and also in the police responsiveness to such crimes (Iyer et al. 2011). This occurred even though *Pradhans* have no jurisdiction over these matters. The increased reporting by women of crimes against them appears instead to reflect a change in their perceptions of the costs – psychic and otherwise – of reporting the crimes.

Is the success of political affirmative action for women generalizable to other socially excluded groups? The 1992 amendment to the constitution of India also made it mandatory to choose a member of a low caste (a former untouchable caste) as *Pradhan* in a fraction of the villages of each state. It might be that the elevation of members of a stigmatized group to a position of power will change the exclusionary mental models *only* if there is some indirect tie or

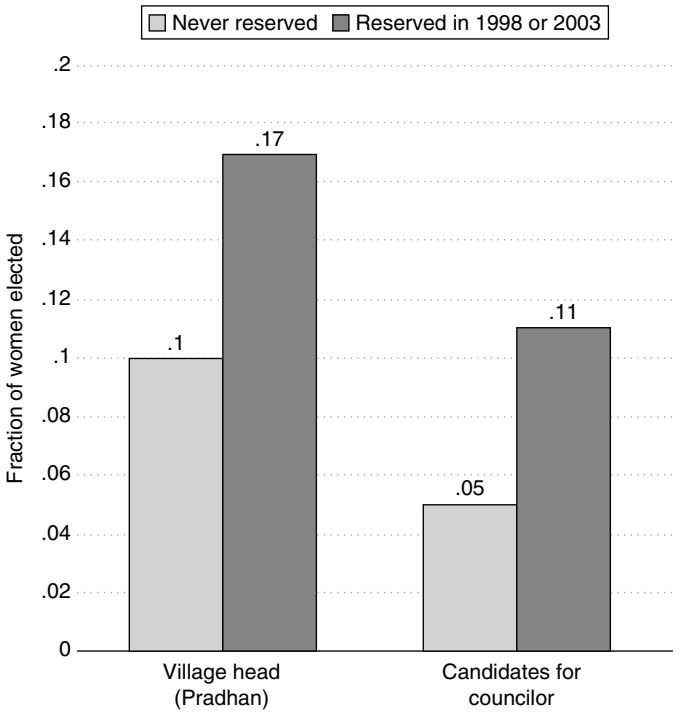


Figure 6.2 The fraction of women elected in free elections
Source: Author; Data Source: Beaman and others (2009).

analogy that provides a basis for a new representation. Such analogies might have been available for women in India: They have roles of pre-eminence as mothers-in-law and as goddesses. In contrast, high-caste individuals are not likely to encounter any obvious source of analogies to roles of pre-eminence for low castes. The next section focuses on behaviors in public schools that became worse as a result of political reservations for the low caste as village *Pradhans*.

Case study of a failure (to date): The effect of political affirmative action for former untouchable castes in north India

For thousands of years, the low castes of India – historically called untouchables and today called Scheduled Castes (SCs) – were institutionally excluded from basic social, political, and economic rights. They were denied access to temples and schools, forced to live in segregated quarters of the villages, and denied the right to own assets and enter most occupations. In some parts of India, untouchables could not even walk through higher-caste neighborhoods. After the independence of India, an end was put to the legal recognition of

the rules of caste, but the social behavior of high castes is today still largely governed by the norms of the caste system.

What can one expect from political affirmative action in village governments for the low castes? The research findings discussed next, based on Pandey (2005, 2010), are from a different state of India than the research findings, discussed above, on political affirmative action for women. The two sets of findings also use different outcome measures. Thus the findings are not directly comparable. Nonetheless, they suggest that political affirmative action for a stigmatized group may not mitigate its social exclusion but instead may actually lower the quality of the public services that the group depends on.

The state of Uttar Pradesh had two distinct sets of land tenure institutions under British colonial rule: *elite control areas* (*elite areas*, for short) and *non-elite control areas* (*non-elite areas*, for short). In elite areas, the landlords were responsible for paying the land tax to the British colonial authority. The landlords had a free hand to exercise political, economic, and judicial authority over the villages. In contrast, in non-elite areas, the cultivators were responsible for paying the land tax to the British colonial power, and the authority of the landlords was much more circumscribed. The land tax was abolished after India's independence, but the legacy of elite control is still evident in the greater political presence of individuals from dominant classes in local government. Local governance and public school outcomes are worse in villages situated in regions with a history of elite control.

Pandey's evaluation of the impact of political affirmative action for SCs focuses on the impact on public schools. Most SC families depend on public schools, and most high-caste families do not.[7] Her evaluation occurred in 2002–03 – within the second election term after SC affirmative action came into effect. In this period, oversight and control over village schools began to be devolved to village governments. In many random visits to schools, Pandey collected data on teacher effort and attendance at school and also on student learning outcomes. By all these measures, SC reservations lowered the quality of the public schools. Teacher effort and the performance of students declined, fees exacted by teachers increased, and less stipend money reached SC students. The negative effects were significant mainly in non-elite areas. In these areas, reservations significantly increased excess fees charged per student, from 22 to 30 rupees – see Figure 6.3. In contrast, in elite areas, where excess fees were already 30 percent greater than in non-elite areas, reservations made no change in the level of excess fees charged per student (about 40 rupees, equivalent to one U.S. dollar and close to the daily wage for unskilled labor).

Figure 6.4 reports the scholarships paid to SC students in SC-reserved and non-reserved villages. The level of the scholarship is set by the state government, but the village teachers administer it. Like Figure 6.3, Figure 6.4 shows that in elite areas without the reservation, teachers were more corrupt than in

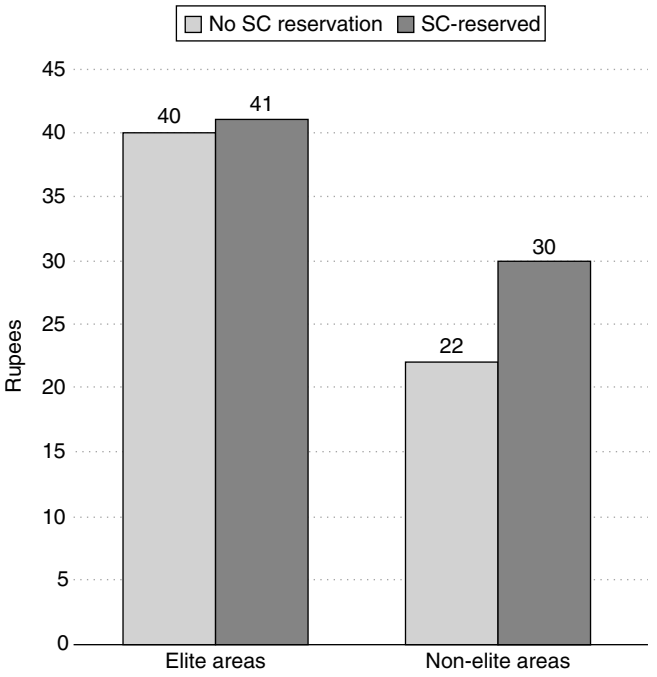


Figure 6.3 Excess fees charged to students
Note: Excess fees are the difference between average fees paid by students and the legally required fees.
Source: Author; Data Source: Pandey (2005).

non-elite areas. Scholarships paid to SC students in elite areas were only 70 percent of the level paid in non-elite areas. In elite areas, reservation again had no effect. But as before, reservations worsened corruption in non-elite areas. Scholarship payments fell on average by 23 percent in reserved, non-elite areas.

Figure 6.5 considers the fraction of teachers who actively teach during random visits to the schools. Again the significant effects occurred only in non-elite areas. For simplicity, the figure shows only the results for the non-elite areas and distinguishes SC from non-SC teachers. Reservations reduced teacher effort for non-SC teachers: in non-reserved areas, they were active 65 percent of the time; but in reserved areas, they were active only 50 percent of the time. In contrast, reservation increased effort by SC teachers from 66 percent to 79 percent. Regression results, not shown here, indicate lower test scores in SC-reserved villages compared to non-SC reserved villages in non-elite areas. There is no difference associated with reservation in elite areas.

To summarize, village governments and public schools are dysfunctional to a greater degree in elite than in non-elite areas. SC reservation is not associated with worse outcomes in elite areas, but neither is it able to reverse the poor

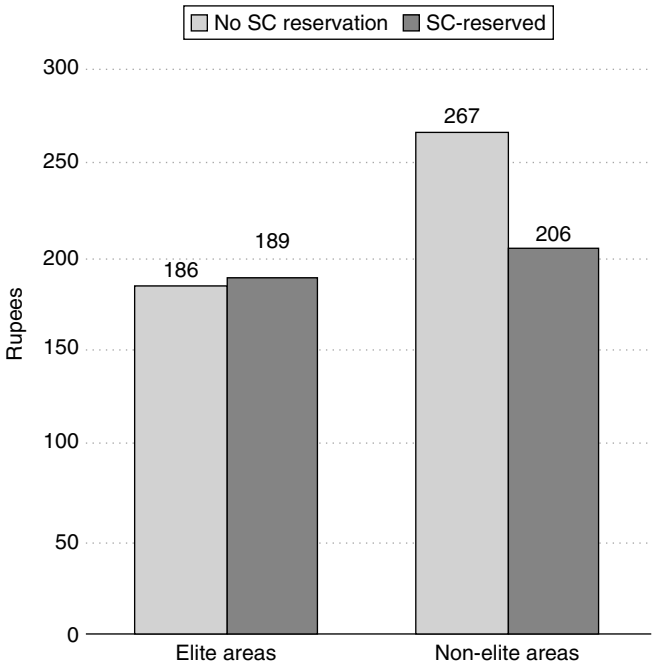


Figure 6.4 Scholarships paid to SC students
Source: Author; Data Source: Pandey (2005).

quality of outcomes. In the non-elite areas, SC-reserved villages compared to non-reserved villages end up with worse schools and greater corruption. Worse governance is possibly due to intimidation of the SCs and their economic dependence on the high castes. With minimal land ownership, SCs commonly depend on high-caste landed households in the village for employment (Lieten and Srivastava 1999). Sixty percent of low-caste *Pradhans* report facing physical violence, threats, and manipulation of votes during their elections and believe that other village council members did not cooperate with them.

Teacher absenteeism is facilitated by a nexus between local elites and teachers who share common caste and class backgrounds. A large proportion of teachers come from the high castes and own land. In the sample, 83 percent of the teachers are of castes higher than SC and only 17 percent are SC. The average land area owned by a teacher is 2.44 acres, compared to only 1.81 acres for a low-caste village council head.

There is no direct evidence on perceptions to compare to the evidence in the preceding discussion of affirmative action for women. But the evidence on governance provides no reason to believe that, at least in the short run, exposure to "powerful" SCs – that is, to SC *Pradhans* – changed the perceptions

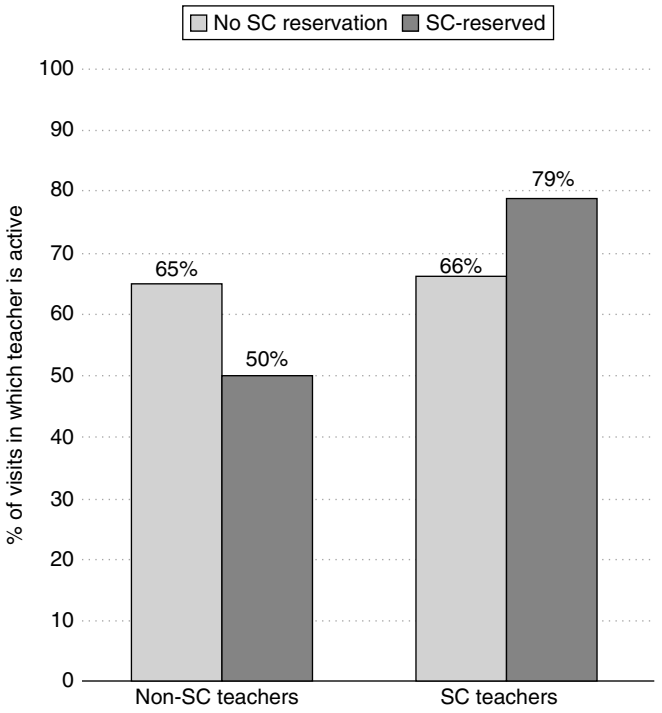


Figure 6.5 Teacher effort in non-elite areas
Source: Author; Data Source: Pandey (2005).

that high-caste individuals held of SCs. The increase in effort by SC teachers in reserved villages is one hopeful sign, but there is insufficient evidence to explain why it occurred.[8]

Intervention 3. Changing the salience of social identities

For interpreting a situation, individuals may have multiple mental models, including multiple social identities. As noted above, the models are not necessarily consistent. Seemingly irrelevant features of the social context may determine which mental model an individual uses. As a result, changes in context that would have no influence in the rational actor model can be effective interventions in practice.

Two broad factors determine which of several social identities that apply to an individual is activated in a context: (1) his perception of the situation in which he finds himself, and (2) the relevance he attributes to the particular social identity as a factor in the situation (Okamura 1981). The pioneering

studies of the effect on performance of priming identity find that merely checking a box to indicate race before taking a standardized test lowers the performance of African-Americans but has no effect on the performance of whites (Steele and Aronson 1995, 1998). The "race prime" appears to raise the consciousness of negative stereotypes among African-American students, and "stereotype threat" reduces performance: "Participants who experience stereotype threat spend more time doing fewer items less accurately" (Steele and Aronson 423).

Lab-in-the field experiments: Social identity primes in India and China

A feature of the Indian caste system that makes it well-suited to identifying the effect of social identity is that caste is fixed by birth and the traditional meaning of caste categories is not in doubt: High-caste individuals are considered socially and intellectually superior in all respects to low-caste individuals (SCs). However, evidence of a new social order is today visible to every schoolchild in the measures to encourage SC parents to enroll their children in school and to participate in the political process. Yet SC children are still likely to encounter the traditional order of caste and untouchability in their own experiences, in the fables they learn, and in the continued discrimination, insults, and atrocities against upwardly mobile members of low castes.

In field experiments, Hoff and Pandey (2006, 2014) assessed the effect on intellectual performance of two manipulations: (1) making caste identity public in groups consisting of three high-caste and three low-caste children, and (2) making caste identity public in groups consisting of six children who were only low caste or only high caste. Segregation of high and low castes is an obvious mark of civic privileges and disabilities (Jodhka 2002). Participants in the experiment were junior high school boys. The participants were asked to solve mazes under monetary incentives.

Participants were randomly assigned to one of the three conditions: condition (1), above, called Revealed Mixed; condition (2), above, called Revealed Segregated; or the control condition, which included three high-caste and three low-caste boys in a session but did not make caste identity public. Thus the control condition is called Caste Not Revealed. Since, in general, the children came from six different villages, their caste would not be known to the other children in the session.

The control condition showed that low-caste boys solve mazes just as well as high-caste boys; see Figure 6.6. However, publicly revealing caste in mixed-caste groups created a significant, 23 percent caste gap in total mazes solved in favor of the high castes, controlling for other individual variables. The low-caste boys may have felt "I can't or don't dare to excel." Publicly revealing caste in caste-segregated classrooms – which is a marker of high-caste entitlement – depressed the performance of both high-caste and low-caste boys, and again their

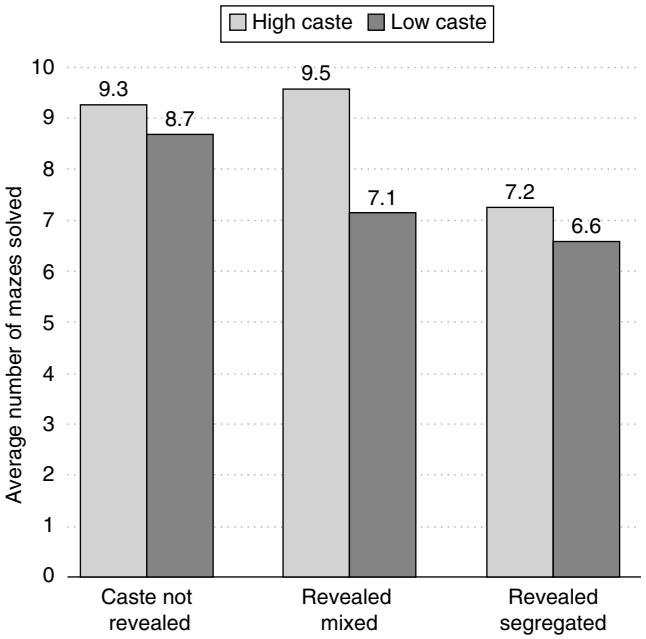


Figure 6.6 The performance of high- and low-caste boys in India
Note: The bars show the sum of mazes solved in two 15-minute rounds under piece-rate incentives. The caste gap in performance is statistically significant with 95 percent confidence in the Revealed Mixed treatment.
Source: Author; Data Source: Hoff and Pandey (2014).

performance was statistically indistinguishable. If segregation evokes a sense of entitlement to the high caste, the high-caste boys may have felt, "Why try?"

The experiment on revealing a stigmatized identity in a mixed group was replicated in China for both boys and girls (Afridi, Li, and Ren, 2015), although in this experiment the identity treatment was stronger: In the identity treatment, students completed a pre-experiment survey that asked questions about their own social identity and about the characteristics of groups with their own and other social identities. The experimental subjects in China were elementary school children in grades 3–6 (ages 8–12), drawn from two social categories: (1) those from households legally classified as urban Beijing households, which is a socially privileged category; and (2) those from households legally classified as rural non-Beijing, which is a socially disadvantaged category. The household registration system in China, known as *hukou*, classifies citizens based on the birthplace of either their parents or grandparents. Those categorized as urban residents of the city in which they live are favored in housing, jobs, access to schools, and public benefits. Those who are categorized as rural migrants are disfavored in all these respects. Unlike categories of gender, caste,

or class, the categories under the household registration system are a transparent administrative creation. However, they have been reinforced through four decades of differential treatment under Chinese law. Will cueing the categories in the classroom affect students' performance?

Just as in the study in India, in the study in China there was no significant gap between performance under piece-rate incentives of the high- and low-status groups in the control treatment, in which identity was not revealed. But making *hukou* identity salient significantly reduced the performance under piece rate incentives of the low-status category, both absolutely and relative to that of the high-status category. Figure 6.7 shows that when *hukou* identity was made salient, the performance (pooling boys and girls) of the low-status students declined by 13 percent, whereas the performance of the high-status students rose by 8 percent.

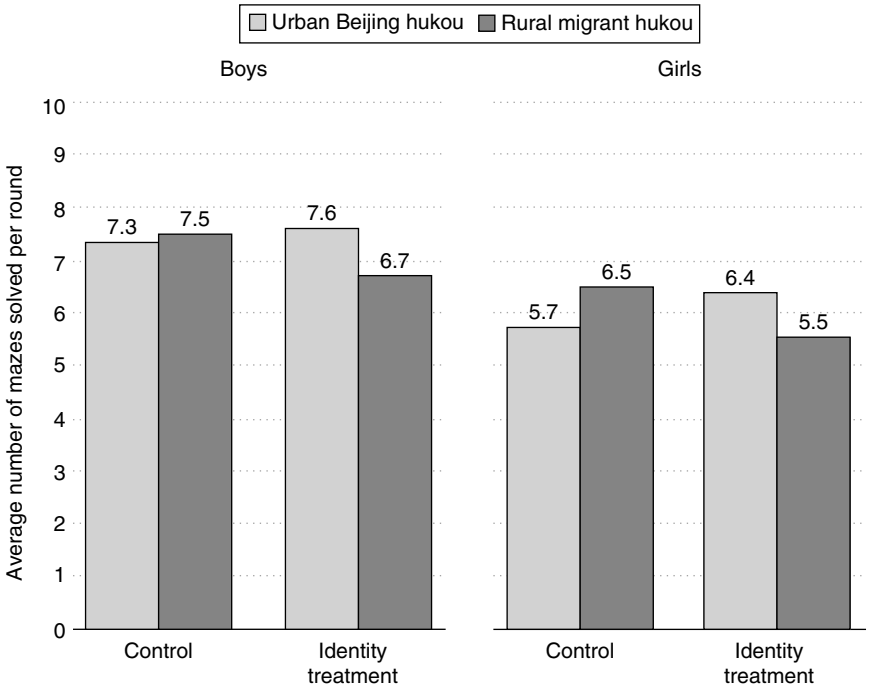


Figure 6.7 The performance of boys and girls with non-migrant and migrant status in Beijing

Note: The data report the average number of mazes solved in each round of 15 minutes under piece-rate incentives. The decline in performance in the Identity treatment for those with rural migrant *hukou* is marginally significant for boys, $p<10\%$, and significant for girls, $p<5\%$. The increase in performance in the Identity treatment for those with urban Beijing *hukou* is not significant when boys' and girls' scores are analyzed separately, but when the results by gender are pooled, the increase in performance is marginally significant, $p<10\%$.

Source: Author; Data Source: Afridi, Li, and Ren (2015).

The examples from India and China illustrate the power of social constructions that rank social groups – such as caste and *hukou* – to "make up people." Even when performance entails uniform incentives across groups, performance can depend sensitively on cues to social identity that would be irrelevant under the rational actor model. The cues can shift performance and ability in ways that mirror the socially constructed ranking of the groups. This is true even for administratively created social identities in an ethnically homogeneous society such as China.

Conclusion

Beliefs play a larger role in behavioral economics than in standard economics, which is based on the rational actor model. Mental models based on pre-existing beliefs about the way the world is organized structure how people process information. The models shape what an individual pays attention to, how he perceives a situation and construes its meaning, what he remembers, and how self-confident he is. Institutions that historically excluded certain groups on the basis of ascriptive characteristics can thus have persistent effects on mental models and thereby reproduce social exclusion even after the formal barriers have been abolished. Mental models can be powerful and should be targets of policy intervention when they contribute to social exclusion.

This paper discussed three kinds of policy. Intervention 1 – group deliberation – contributed to the abandonment of female genital cutting in Senegal. Intervention 2 – political affirmative action – improved outcomes for women in rural India but did not do so for the low castes, at least not in the short run by the available measures. In public schools, on which the low castes disproportionately depend, the reservation of village political leader for a low-caste individual actually increased high-caste teacher absenteeism and lowered student learning. Intervention 3 – reducing the salience of a stigmatized identity – eliminated a performance gap between students in the stigmatized and dominant groups.

A behavioral intervention not discussed in this paper are programs that mix rich and poor children. This intervention has had some success. In India, when pre-school children at the 95th percentile of income were mixed in schools with children at the 25th percentile, the rich children became more pro-social and generous towards other children (Rao 2013). Personal interactions in school between rich and poor children also caused the rich children to discriminate less against poor children (measured by their choice of teammates in an incentivized sports contest) and to socialize more with poor children outside school.

This paper draws on insights from psychology, sociology, anthropology, and neuroscience that are at odds with the rational actor model, but this is not to

imply that the rational actor model does not remain of central importance in economics. In general, behavioral and rational actor perspectives complement each other. By highlighting the diversity of factors that perpetuate social exclusion, behavioral economics broadens the set of interventions to reduce it. Increasingly, welfare programs try to address both structural and behavioral factors in social exclusion.[9]

Notes

1. Surveys are Nisbett and Norenzayan (2002) and Camerer, Loewenstein, and Prelec (2005). Demeritt and Hoff (2015) consider the applications to behavioral development economics.
2. For similar findings, see Lane et al. (2007).
3. This point is related to earlier work in the rational actor model that showed how institutions and beliefs about the world can be jointly determined. For instance, Piketty (1995) showed that the belief that earnings largely reflect effort, not luck, affects political institutions (the welfare state), which can then lead to choices (low work effort) such that the belief is never refuted even if it is false. The focus here is on the joint determination of *mental models* (uncritically held and taken-for-granted beliefs) and *institutions* – in particular, institutions of social exclusion. Beliefs that are part of mental models can be difficult to overturn even when they are false, and play a larger role in behavioral economics than in the rational actor model. In the former but not the latter, beliefs influence information processing and can produce "equilibrium fictions" – a world created and sustained by schematic cognition (Hoff and Stiglitz 2010).
4. Notable exceptions are Greif (2005) and North (2005).
5. Rabin and Schrag (1999) present a model of confirmation bias in which individuals with strong prior beliefs may *never* update their beliefs.
6. It is not always true that a community practices FGC almost universally or not at all. In the Sudan, Efferson et al. (2015) find large variation across villages in the extent of FGC; it ranges continuously from very few women to almost every woman. The determinants of FGC in the Sudan are under investigation.
7. In 2002–2003, 42 percent of children in public schools were SC, while only 15 percent were high caste, even though the shares of each caste group in the population were about the same (24 percent for SC and 27 percent for high caste in the sample villages).
8. Jensenius (2015) finds no impact on development or redistribution to SCs of political reservations for SCs in 15 Indian states.
9. See, for example, Carneiro, Galasso, and Ginja (2015).

References

Afridi, Farzana, Sherry X. Li, and Yufei Ren (2015) "Social Identity and Inequality: The Impact of China's *Hukou* System," *Journal of Public Economics*, vol. 123, pp. 17–29.

Axelrod, Robert (1973) "Schema Theory: An Information Processing Model of Perception and Cognition," *American Political Science Review*, vol. 67, no. 4, pp. 1248–466.

Beaman, Lori, Raghabendra Chattopadhyay, Esther Duflo, Rohini Pande, and Petia Topalova (2009) "Powerful Women: Does Exposure Reduce Bias?," *The Quarterly Journal of Economics*, vol. 124, no. 4, pp. 1497–540.

Beaman, Lori, Esther Duflo, Rohini Pande, and Petia Topalova (2012) "Female Leadership Raises Aspirations and Educational Attainment for Girls: A Policy Experiment in India," *Science*, vol. 335, no. 6068, pp. 582–6.

Bertrand, Marianne, and Sendhil Mullainathan (2004) "Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination," *American Economic Review*, vol 94, no. 4, pp. 991–1013.

Bruner, Jerome (1990) *Acts of Meaning*. Cambridge, MA: Harvard University Press.

Camerer, Colin, George Loewenstein, and Drazen Prelec (2005) "Neuroeconomics: How Neuroscience Can Inform Economics," *Journal of Economic Literature*, vol. XLIII, pp. 9–64.

Carneiro, Pedro, Emanuela Galasso, and Rita Ginja (2015) "Tackling Social Exclusion: Evidence from Chile," World Bank, Policy Research Working Paper 7180.

Cislaghi, Beniamino, Dianne Gillespie, and Gerry Mackie (2014) "Values Deliberations and Collective Action in Rural Senegal." Manuscript. University of California, San Diego.

Demeritt, Allison and Karla Hoff (2015). "'Small Miracles'—Behavioral Insights to Improve Development Policy: *The World Development Report 2015*." In Timothy Besley, ed. *Contemporary Issues in Development Economics*, International Economic Association, London: Palgrave Macmillan.

Denzau, Arthur T. and Douglass C. North (1994) "Shared Mental Models: Ideologies and Institutions," *Kyklos*, vol. 47, no. 1, pp. 3–31.

DiMaggio, Paul (1997) "Culture and Cognition." *Annual Review of Sociology*, vol. 23, no. 1, pp. 263–87.

Douglas, Mary (1986) *How Institutions Think*. Syracuse, NY: Syracuse University Press.

Efferson, Charles, Sonja Vogt, Amy Elhadi, Hilal El Fadil Ahmed, and Ernst Fehr (2015) "Female Genital Cutting Is Not a Social Coordination Norm," *Science*, vol. 349, no. 6255, pp. 1446–7.

Elkins, David J. and Richard E.B. Simeon (1979) "A Cause in Search of Its Effect, or What Does Political Culture Explain?," *Comparative Politics*, vol. 11, no. 2, pp. 127–45.

Friedland, Roger and Robert R. Alford (1991) "Bringing Society Back In: Symbols, Practices and Institutional Contradictions," *The New Institutionalism in Organizational Analysis*, October 25, pp. 232–63.

Gennaioli, Nicola and Andrei Shleifer (2010) "What Comes to Mind," *The Quarterly Journal of Economics*, vol. 125, no. 4, pp. 1399–433.

Glaeser, Edward L, Bruce Sacerdote, and Jose A. Scheinkman (1996) "Crime and Social Interactions," *The Quarterly Journal of Economics*, vol. 111, no. 2, pp. 507–48.

Greif, Avner (2005) "Cultural Beliefs and the Organization of Society: A Historical and Theoretical Refleciton on Collectivist and Individualist Societies," *Journal of Political Economy*, vol. 102, no. 5, pp. 912–50.

Hicks, Esther K. (1993) *Infibulation: Female Mutilation in Islamic Northeastern Africa*. New Brunswick, NJ: Transaction Books.

Hoff, Karla, and Priyanka Pandey (2006) "Discrimination, Social Identity, and Durable Inequalities," *American Economic Review*, vol. 96, no. 2, pp. 206–11.

Hoff, Karla, and Priyanka Pandey (2014) "Making Up People – The Effect of Identity on Performance in a Modernizing Society," *Journal of Development Economics*, vol. 106, pp. 118–31.

Hoff, Karla, and Joseph Stiglitz (2010) "Equilibrium Fictions: A Cognitive Approach to Societal Rigidity," *American Economic* Review, vol. 100 (May), pp. 141–6.

Iyer, Lakshmi, Anandi Mani, Prachi Mishra, and Petia Topalova (2011) "The Power of Political Voice: Women's Political Representation and Crime in India," *American Economic Journal: Applied Economics*, vol. 4, no. 4, pp. 165–93.

Jensenius, Francesca (2015) "Development from Representation? A Study of Quotas for Lower Caste Groups in India," *American Economic Journal Applied Economics*, vol. 7, no. 3, pp. 196–220.

Jodhka, Surinder S. (2002) "Caste and Untouchability in Rural Punjab," *Economic and Political Weekly*, vol. 37, no. 19, pp. 1813–23.

Kahneman, Daniel (2003) "Maps of Bounded Rationality: Psychology for Behavioral Economics," *American Economic Review*, vol. 93, no. 5, pp. 1449–75.

Kahneman, Daniel (2011) *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.

Kitayama, Shinobu and Jiyoung Park (2010) "Cultural Neuroscience of the Self: Understanding the Social Grounding of the Brain," *Social Cognitive and Affective Neurosciene*, vol. 5, pp. 111–29.

Kyvig, David (1979) *Repealing National Prohibition*. Chicago: University of Chicago Press.

Lane, K. A., Mahzarin Banaji, B.A. Nosek, and A.G.Greenwald (2007) "Understanding and Using the Implicit Association Test: IV: Procedures and Validity," in B. Wittenbrink and N. Schwarz (eds), *Implicit Measures of Attitudes: Procedures and Controversies*. New York: Guilford Press, pp. 59–102.

Lieten, Georges Kristoffel, and Ravi Srivastava (1999) *Unequal Partners: Power Relations, Devolution and Development in Uttar Pradesh*. New Delhi: Sage Publications India Pvt Ltd..

Loury, Glenn C. (2002) *The Anatomy of Racial* Inequality. Cambridge, MA: Harvard University Press.

Mackie, Gerry (2003) "Female Genital Cutting: A Harmless Practice?," *Medical Anthropology Quarterly*, vol. 17, no. 2, pp. 135–58.

Mackie, Gerry, and John LeJeune (2009) "Social Dynamics of Abandonment of Harmful Practices: A New Look at the Theory," Special Series on Social Norms and Harmful Practices, Innocenti Working Paper 2009–06, UNICEF Innocenti Research Centre, Florence.

Mercier, Hugo, and Dan Sperber (2011) "Why Do Humans Reason? Arguments for an Argumentative Theory," *Behavioral and Brain Sciences*, vol. 34, no. 2, pp. 57–74.

Mitchell, J.P, B.A. Nosek, and Mahzarin R Banaji (2003) "Contextual Variations in Implicit Evaluation," *Journal of Experimental Psychology:* General, vol. 132, no. 3, pp. 455–69.

Nisbett, Richard E. and Ara Norenzayan (2002) "Culture and Cognition," in *Stevens' Handbook of Experimental Psychology*, third edition, vol. 2. New York: John Wiley & Sons, pp. 561–98.

North, Douglass C. (2005) *Understanding the Process of Economic Change*. Princeton, NJ: Princeton University Press.

Okamura, Jonathan Y. (1981) "Situational Ethnicity," *Ethnic and Racial Studies*, vol. 4, no. 4, pp. 452–65.

Pager, Devah, Bruce Western, and Bart Bonikowski (2009) "Discrimination in a Low-Wage Labor Market: A Field Experiment," *American Sociological Review*, vol. 74, no. 5, pp. 777–99.

Pandey, Priyanka (2005) "Service Delivery and Corruption in Public Services: How Does History Matter and Can Mandated Political Representation Reverse the Effects of History?," manuscript.

Pandey, Priyanka (2010) "Service Delivery and Corruption in Public Services: How Does History Matter?," *American Economic Journal: Applied Economics*, vol. 2, no. 3, pp. 190–204.

Piketty, Thomas (1995) "Social Mobility and Redistributive Politics," *The Quarterly Journal of Economics*, vol. 110, no. 3, pp. 551–84.

Rabin, Matthew, and Joel L. Schrag (1999) "First Impressions Matter: A Model of Confirmatory Bias," *The Quarterly Journal of Economics*, vol. 114, no. 1, pp. 37–82.

Rao, Gautam (2013) "Familiarity Does Not Breed Contempt: Diversity, Discrimination and Generosity in Delhi Schools," Working Paper, Department of Economics, Harvard University.

Ross, Lee, and Richard E. Nisbett (1991) *The Person and the Situation: Perspectives of Social Psychology*. New York: McGraw-Hill.

Shleifer, Andrei (2012) "Psychologists at the Gate: A Review of Daniel Kahneman's *Thinking, Fast and Slow*," *Journal of Economic Literature*, vol. 50, no. 4, pp. 1–12.

Steele, Claude M., and Joshua Aronson (1995) "Stereotype Threat and the Intellectual Test Performance of African Americans," *Journal of Personality and Social Psychology*, vol. 69, no. 5, pp. 797–811.

Steele, Claude M., and Joshua Aronson (1998) "Stereotype Threat and the Test Performance of Academically Successful African Americans," *The Black-White Test Score Gap,* eds. Christopher Jencks and Meredith Phillips. Washington, DC, Brookings Instituiton, pp. 401–27.

Strauss, Claudia and Naomi Quinn (1997) *A Cognitive Theory of Cultural Meaning*. Cambridge, UK: Cambridge University Press.

Sunstein, Cass R. (1996) "Social Norms and Social Roles," *Columbia Law Review*, vol. 96, no. 4, pp. 903–68.

Swidler, Ann (1986) "Culture in Action: Symbols and Strategies," *American Sociological Review*, vol. 51, no. 2, pp. 273–86.

Todd, Peter M. and Gerd Gigerenzer (2000) "Précis of Simple Heuristics That Make Us Smart," *Behavioral and Brain Sciences*, vol. 23, no. 5, pp. 727–41.

World Bank (2013) *Inclusion Matters: The Foundation for Shared Prosperity*. Washington, DC: World Bank.

World Bank Group (2015) *World Development Report 2015: Mind, Society, and Behavior*. Washington, DC: World Bank.

Origin and Evolution of Cognitive Frames: Comments on "Behavioral Economics and Social Exclusion: Can Interventions Overcome Prejudice?" by Karla Hoff

Paola Giuliano
UCLA Anderson School of Management

1 Introduction

One of the oldest, most fundamental questions in economics concerns the elimination of social exclusion. Standard economic interpretations of social exclusion assume rationality and consider the behavior of the economically disadvantaged as calculated adaptations to prevailing circumstances. Policy interventions can therefore ameliorate these adverse circumstances by improving access to financial services, bolstering labor markets, and so on.

Hoff (2014) posits that behavioral economics can provide a larger set of policy tools to improve social exclusion than more traditional economic models. Hoff's behavioral perspective draws from lessons in social and cognitive psychology and proposes that individuals use "cognitive frames" when making decisions. Policies can therefore be very helpful in changing how to think and in modifying acquired cognitive frames.

The idea that individual behavior is driven by cognitive frames is present in many disciplines. For example, in cultural evolution, Boyd and Richerson (1985, 2005) construct models in which individuals make decisions using decision-making heuristics or rules of thumb. These decision-making heuristics have evolved given our need to make decisions in a complex and uncertain environment. The main idea is that if information is either costly or imperfect, the use of "rules of thumb" in decision making can arise optimally. However, by relying on these "rules of thumb," individuals may not behave in a manner that is optimal in every instance.

Decision-making heuristics may manifest themselves in a number of ways, for example, they can take the form of emotions or gut feelings about what the right or wrong action is in a particular situation. These emotions may range from deeply held beliefs about the extent to which others can be trusted, whether women should work outside the home, and whether it is important to punish those who have cheated on others in the community.

Though we have yet to fully understand the precise mechanics underlying our reliance on emotions, mental constructs, and cognitive frames, ample evidence shows that these shortcuts exist, and that they do influence decision making (De Sousa, 1987; Damasio, 1994; Elster, 1999; Gigerenzer, 2007).

2 How policies can modify cognitive frames

Hoff takes as a given the relevance of cognitive frames and provides three examples of how they could be successfully modified: group deliberation, political affirmative action, and priming and performance.

2.1 Group deliberation

People tend to be irrationally loyal to their own beliefs for various reasons, such as the existence of a confirmation bias or the presence of false consensus.[1] Group deliberation could help change mental frames and limit various cognitive biases.

Several questions arise regarding the validity of group deliberation as a potential policy tool for modifying mental frames. In the case study Hoff describes, on the abandonment of female genital mutilation in Senegal, it's unclear exactly how group deliberation tipped the societal equilibrium. Was it through the way information was provided initially? Hoff recounts that a trained facilitator led discussions in each group for two to three years. Did these facilitators intervene outside women's groups, for example, to change men's belief that only women with female genital cutting should be married? Understanding how group deliberation works is key to making the experience in Senegal replicable in other environments.

A second question is whether cognitive frames reappear any time an individual faces a new pool of people. Were the groups of women fixed over time? Was each group's facilitator the same over time, or did different facilitators intervene during the case-study period?

The final consideration regarding group deliberation, from a policy perspective, is that although it can change individual mental frames, it's only a first (albeit necessary) step in triggering the change in values and the way of thinking regarding female genital cutting. Coordination throughout society is the most important component in generating complete change in cognitive frames. In Hoff's example, the elimination of female genital cutting would not have been possible without a change in the overall societal equilibrium.

Cognitive frames, like social norms, typically emerge in environments characterized by multiple equilibria, to keep the community in a preferred equilibrium (Kandori, 1992). The practice of female genital mutilation, therefore, could be seen as a societal equilibrium: it can be traced back to the second century BC, it is transmitted by women across generations, it promotes

marriage and family honor, and it enhances the status of women. In multiple-equilibria models, two elements are necessary to switch the equilibrium: a change in environment (in this case, group deliberation provided new information) and a coordinated individual reaction to the change. The description of the Senegal case provides important information about the first step, but we know nothing about the coordinated individual reaction to the change (for example, how men changed their minds) or on how women outside the groups changed their minds (unless everybody was treated).

Another interesting case helps us understand the nature of the problem. In 1978, the International Centre for Diarrhoeal Disease Research, in Bangladesh, launched its Maternal Child Health and Family Planning project, covering 70 villages in Matlab Thana, in the Comilla district. All households were visited by a community health worker once every two weeks, and contraceptives were provided to them free of costs. In rural Bangladesh, the traditional norm was characterized by early, universal marriage, followed by immediate, continual child-bearing. Religious authorities legitimized the norm and enforced the rules that sustained the equilibrium. The sudden, unexpected availability of modern contraceptives through this program opened up the possibility for new equilibria, in which a sizeable fraction of women in the village began to regulate fertility, ignoring the religious sanctions. The critical point was social uncertainty: no woman knew what level of contraceptive prevalence could ultimately be sustained in her community.

Munshi and Myaux (2006) studied this case, looking at how individuals learned via social interactions about the new reproductive equilibrium that emerged in their communities following the introduction of the contraceptives. Consistent with the view that morphing social norms are driving changes in reproductive behavior in these communities, the authors find that the individual's contraception decisions responded strongly to changes in contraceptive prevalence in her own religious group within the village, whereas cross-religious effects were entirely absent.

This example illustrates that conformity to the norm coming from religious groups was essential to the equilibrium change. It would be interesting to know if and how group deliberation was more powerful than any other mechanism adopted in the past regarding the release of information on female genital mutilation, and also how the new equilibrium in the community was reached.

2.2 Political affirmative action

Institutions can teach people how to "see" and how to think; therefore, policies that change institutions have the potential to change mental frames. A big question mark for the implementation of this recommendation is that the probability of accepting different institutions varies. In a provocative book,

Emmanuel Todd (1990) claims that we tend to adopt institutions and ideologies that are compatible with beliefs or models we learned from our parents. Bisin and Verdier (2014) similarly claim that the successful imposition of institutions depends on whether the underlying beliefs in societies are compatible with the institutions imposed.[2] Finally, if beliefs, values, and cognitive frames have deep historical roots (Alesina, Giuliano, and Nunn, 2013) and tend to be "sticky," institutions might not be very successful in changing them.

Hoff's two examples could be broadly in line with the above-mentioned views. She describes two policy changes – one successful and one not – that aimed to limit the social exclusion of women and scheduled castes. An Indian constitutional amendment that made it mandatory to reserve in a fraction of villages the position of village head for women was very successful, according to Hoff, because women have pre-eminent roles in other aspects of life in India, and therefore it was easy to extend this mental frame to other domains. This mental association was not possible, however, for scheduled castes; thus, a similar type of intervention (though evaluated in a different state and that looked at different outcomes) was unsuccessful.

Whereas policies can influence mental frames, they must evince the underlying societal beliefs that could be compatible with specific policy choices. Designing the proper policy is thus a delicate balancing act, requiring a deep knowledge of the underlying historical societal beliefs.

2.3 Priming and performance

Policies reducing the salience of social identity can lessen the sting of social exclusion. In a series of experiments Hoff et al. (2006, 2014) study the relationship between priming of identity and performance by looking at high school boys from low and high castes solving mazes in various treatments. The authors find that low-caste and high-caste boys solve mazes similarly when segregated or during group play when their caste is not revealed. However, when students play in mixed groups and the caste is revealed, the performance gap increases, with low castes performing worse.

Reducing or emphasizing social identity could be a powerful policy instrument. Some difficulties could arise, however. For one thing, it's often unclear whether identities or stereotypes are more powerful. Shih et al. (1999) perform a series of experiments in which an ethnic identity or a gender identity is activated when performing quantitative tasks. Two dominant stereotypes are that Asians have superior quantitative skills compared to other ethnic groups and that women have inferior quantitative skills compared to men. Asian-American women performed better in quantitative tasks when their ethnic identity was activated but worse when their gender identity was activated. The authors conclude that stereotypes could be more powerful than identities in driving behavior.

Also, individuals derive identity from the groups to which they belong. People may regularly alternate among different salient identities. A working mother might think of herself primarily as a mother when in the company of her children, but see herself as a professional while at work. Triggering the correct identity, therefore, will be crucial for implementing successful policies, for example, when promoting banking or social programs.

3 Conclusions

Many intriguing policy possibilities can unfold when we adopt a behavioral economics perspective on social exclusion. Several questions remain open and are worthy of investigation. Given the relevance of cognitive frames and rules of thumb, it's important to ask where the differences in mental frames come from. Recently, Alesina, Giuliano, and Nunn (2013) showed that differences in beliefs about the "proper" role of women in society can go as far back as preindustrial societies. If differences in cognitive frames are so ancient, more research needs to be done on why different groups have different cognitive frames to start with. In addition, if the past is so persistent, what are the mechanisms through which policies can have a long-lasting effect?

Hoff, in her behavioral perspective, starts by assuming that social exclusion could be driven by different mental frames. Mani et al. (2013) provide a different behavioral approach to social exclusion. Social exclusion or poverty, they assert, means coping not only with a shortfall of money but also with a shortfall of cognitive resources. Socially excluded individuals are therefore less able to make decisions not because of inherent traits or cognitive frames, but because the very context of poverty imposes a load and impedes cognitive capacity. The authors use as an example the experience of Indian farmers, whose income varies throughout the year. In particular, they conducted a field study that used quasi-experimental variation in actual wealth. Indian sugarcane farmers receive income annually at harvest time and find it hard to smooth their consumption. As a result, they experience cycles of poverty (poor before harvest and richer afterward). This disparity allowed the authors to compare cognitive capacity for the same farmer when poor (pre-harvest) versus when richer (post-harvest). The authors found that farmers made higher-return investments just after the harvest compared with later in the season. If social exclusion is driven mostly by lower cognitive ability that's hindered by poverty itself, the set of policies to limit it could be very different from the ones suggested by the persistence of mental frames.

Taking a behavioral economics perspective on social exclusion is likely to both enrich and complicate our views of the role of institutions and policies. If the new policies are founded on a better understanding of societal mechanisms, which have so far been ignored, it clearly seems worth trying to implement them.

Notes

1. According to this second mechanism (Ross, Greene, and House, 1977), people think that others are like them when forming their beliefs. This phenomenon has been shown to be very persistent: neither providing additional information about the population of interest, nor warning individuals about the possibility of false consensus, eliminates the effect.
2. According to Todd (1990), the parent–child relations within the family determine attitudes toward liberal or authoritarian ideologies, whereas the egalitarian inheritance practices among siblings lead to egalitarian ideologies. The appeal of differing modern ideologies results from their mirroring the character of various family types; and such ideologies spread only so far as the geographical extent of the family systems with which they have similarities. This, in his view, should partially explain differences in ideologies around the world, including Communism, Nazism, and Anglo-Saxon liberalism. Todd argues that Communism prevailed in societies with communitarian families (authoritarian in the relationship between parents and children, and egalitarian among siblings). This is because individuals were accustomed, within the family, to the same authoritarian system adopted at the level of government institutions. On the contrary, the absolute nuclear family in England was fertile ground for the development of non-egalitarian capitalism, individualism, and market freedom among siblings (a principle of equality is also fundamental for a communist ideology).

References

Alesina, A., Giuliano, P. and N. Nunn (2013) "On the Origin of Gender Roles: Women and the Plough," *The Quarterly Journal of Economics*, vol. 128, no. 2, pp. 469–530.

Bisin, A. and T. Verdier (2014) "On the Joint Evolution of Culture and Institutions," New York University, mimeo.

Boyd. R. and P. J. Richerson (1985) *Culture and the Evolutionary Process*. Chicago: University of Chicago Press.

Boyd, R. and P. J. Richerson (2005) *The Origin and Evolution of Culture*. New York: Oxford University Press.

Damasio, A. (1994) *Descartes' Error: Emotion, Reason and the Human Brain*. New York: G.P. Putnam's Sons.

De Sousa, R. (1987) *The Rationality of Emotion*. Cambridge, MA: MIT Press.

Elster, J. (1999) *Alchemies of the Mind: Rationality and the Emotions*. New York: Cambridge University Press.

Gigerenzer, G. (2007) *Gut Feelings: the Intelligence of the Unconscious*. New York: Penguin.

Hoff, K. (2006) "Discrimination, Social Identity, and Durable Inequalities," *American Economic Review*, vol. 96, no. 2, pp. 206–11.

Hoff, K. (2014) "Making up People – the Effect of Identity on Performance in a Modernizing Society," *Journal of Development Economics*, vol. 106, pp. 118–31.

Kandori, M. (1992) "Social Norms and Community Enforcement," *Review of Economic Studies*, vol. 59, pp. 63–80.

Mani, A., Mullainathan, S., Shafir, E. and J. Zhao (2013) "Poverty Impedes Cognitive Function," *Science*, vol. 341, no. 6149, pp. 976–80.

Munshi, K. and Myaux, J. (2006) "Social Norms and the Fertility Transition," *Journal of Development Economics*, vol. 80, pp. 1–38.

Ross, L., Greene, D. and P. House, (1977) "The False Consensus Phenomenon: An Attributional Bias in Self-Perception and Social Perception Processes," *Journal of Experimental Social Psychology*, vol. 13, no. 3, pp. 279–301.
Shih, M., T. Pittinsky and N. Ambady (1999) "Stereotype Susceptibility: Identity Salience and Shifts in Quantitative Performance," vol. 10, no. 1, 80–3.
Todd, E. (1990) *L'Invention de l'Europe*. Paris: Seuil.

7
The Effects of Fiscal Redistribution*

Michele Battisti
University of Palermo, LUISS and RCEA

Joseph Zeira
Hebrew University of Jerusalem, LUISS, CEPR and RCEA

7.1 Introduction

Every discussion on income distribution and inequality distinguishes between market income, namely income before tax and without transfers, and disposable, or net income, which is after tax and including transfers. Hence, taxation and transfers create a redistribution of income. This redistribution is usually progressive, as direct taxes and subsidies are progressive, and thus it is supposed to reduce inequality, in the transition from market income to disposable income. This paper focuses on measuring the effect of fiscal policy in income redistribution and in reducing inequality. It also examines which type of fiscal policy is most strongly related to the redistribution of income, are they transfer payments? Is it direct taxation? Or is it the overall measure of fiscal policy, namely public expenditures, which are also known as the size of the public sector?

Figure 7.1 presents a simple scatter-diagram of two variables, the size of the public sector and inequality. The horizontal axis measures the share of public expenditures relative to GDP and the vertical axis measures the Gini coefficient of disposable income. There are 83 countries in the sample for the year 2005. Actually, to avoid cyclical variability, we use five-year averages for each variable, namely the variables in Figure 7.1 are averages over the years 2001–05 for both public expenditures and for the Gini coefficients. Figure 7.1 clearly shows that there is a strong negative correlation between these two variables. It means that countries with a larger public sector tend to have lower inequality

*We are grateful to Kaushik Basu, who supported our work on this paper throughout the whole process. The comments of James Foster helped us significantly. Sharon Hadad first introduced us to the correlation between public expenditure and inequality, which spurred our interest in the topic. We received excellent comments from Paola Giuliano, Giuseppe Ragusa, Domenico Giannone and many more. Remaining errors are, of course, all us.

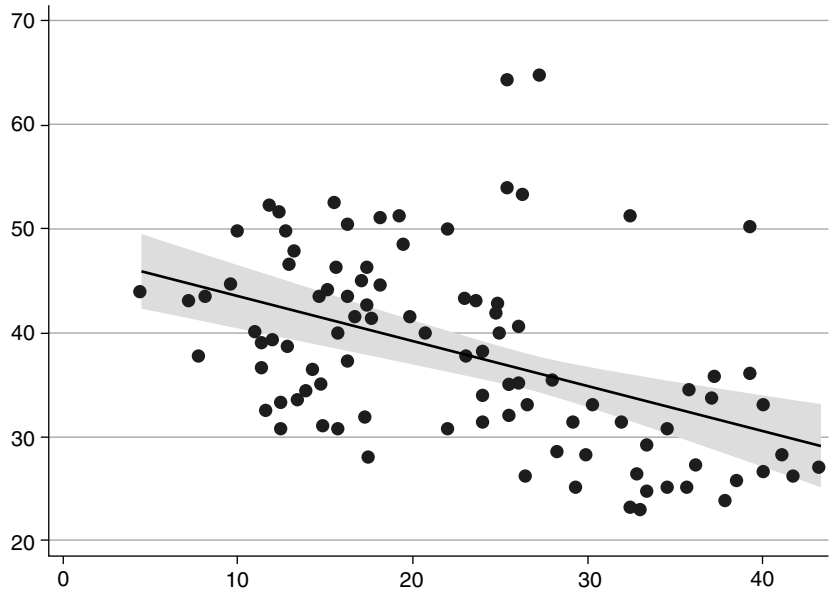


Figure 7.1 Gini of disposable income and percent of public expenditures in GDP across 83 countries in 2005[1]

Note:
[1]Inequality data are from Solt (2009) and public expenditures data are from IMF (2014b).

of disposable income, and vice versa.[1] This observation is our point of departure to a broader empirical investigation. We examine the separate effects on inequality before and after taxes and subsidies, we control for other variables and we test for reverse causality.

In our research we use cross-country regression analysis to estimate the effect of public spending on inequality in disposable income, as measured by the Gini coefficient. In these regressions we control for the Gini coefficient of market income, so that our analysis enables us to measure the specific role of public expenditures in fiscal redistribution. We find that this fiscal variable, public expenditures, captures most of the redistribution done by fiscal policy. We can also quantify its effect and find that every additional percent of GDP to public expenditures reduces the Gini coefficient by 0.4 percentage points. This is a significant effect. We also examine the effect of partial elements of fiscal policy on redistribution, where the main ones are direct taxation and transfers. We find that direct taxation has a comparable effect to fiscal expenditures in reducing inequality and we also find that transfers or subsidies do not have a significant effect on the overall inequality. These results are from a cross-country analysis, but they are reached also when we pool our regressions over time.

We then turn to examine the possibility of reverse causality, namely that fiscal policy is affected by inequality instead of affecting it. Actually, this is a central claim in the literature on political economy, which was raised initially by Meltzer and Richard (1981). They have claimed that if inequality is high, there is more public pressure for redistribution and hence the size of the public sector increases. In order to test for this possibility we run a two-stage regression. In the first stage we test how inequality of market incomes affects fiscal policy, controlling for other variables which should also affect fiscal policy. We then use the resulting estimates of fiscal policy and test how it affects the inequality of disposable income. The results confirm our previous results that causality goes from fiscal policy to disposable income inequality. Actually, our results indicate that the inequality of market incomes has a negative effect on fiscal policy, namely the Meltzer and Richard hypothesis does not pass the empirical test.

Finally, we do not consider only one measure of inequality, namely the Gini coefficient, but we also examine the effect of fiscal redistribution also on the rate of poverty. We find that this rate is also affected by total public expenditures, namely on the overall fiscal policy, but that it is especially affected by transfers. This is unsurprising, since these expenditures are specifically directed toward the poor and should therefore affect their status significantly. Indeed, our cross-country tests show that this effect of transfers on the rate of poverty is quite large.

This paper is related to a few lines of research. The main one is the empirical research on the determinants of inequality. Since data on inequality does not go back too long in time and since data on income distribution is not of high quality for many countries, this literature is not very wide. Barro (2000) is an early example of such a study, which also contains a thorough survey of the previous literature. Barro (2000) focuses mainly on the state of development and its effect on inequality, namely on the Kuznets hypothesis, that inequality increases at the early stages of development and then declines at the later more advanced stages. Barro does not explicitly discuss policy variables in his analysis, while this paper focuses on such variables and especially on fiscal policy. Another important survey of this area of research is Atkinson (1997). Recently there are a number of papers that study the redistribution of income by fiscal policy, from market incomes to disposable incomes, such as Chu, Davoodi and Gupta (2004), IMF (2014a), Niehues (2010), Ospina (2010), Martinez-Vazquez, Vulovic and Moreno-Dodson (2012), Muinello-Gallo and Roca-Sagalés (2013) and Woo et al. (2013). Most closely to our paper are IMF (2014a) and Muinelo-Gallo and Roca-Sagalés (2013) (2013), which examine the relationship between redistribution, fiscal policy and economic growth. Our paper differs from most of these papers in three main ways. First, we do not measure redistribution by the simple difference between the Gini of disposable income and the Gini of

market income, but we rather regress Gini of net income over Gini of market income to get a better estimate to its effect than 1. The second deviation of our paper from these papers is our dealing with the possibility of reverse causality.[2] Third, we measure the effect of fiscal redistribution not only on the overall Gini coefficient but also on income distribution at the bottom, namely on the rate of poverty.

The chapter is structured as follows. Section 7.2 presents the data and the variables used while section 7.3 presents the empirical model. Section 7.4 presents cross-country analysis of the effect of redistribution, while section 7.5 extends the analysis for more periods of time. Section 7.6 examines the possibility of reverse causality. Section 7.7 tests the effect of fiscal redistribution on the rate of poverty and section 7.8 summarizes.

7.2 Variables and data

The main variables used in the paper are the measures of inequality and of fiscal policy. Inequality is measured mainly by the Gini coefficient in percentages. We use Solt (2009) standardized version of the WIID 2008 (UNU-WIDER, 2008), for the two variables of Gini of disposable income and Gini of market income. The Gini of disposable income, or net income, in country j in year t is denoted by GN(j, t), while the Gini of market income in country j in year t is denoted by GM(j, t). In most of our regressions we use averages of the variable over the last five years. Hence, the average variable for period t is defined and denoted in the following way:

$$GN_5(j, t) = [GN(j, t-4) + GN(j, t-3) + GN(j, t-2) + GN(j, t-1) + GN(j, t)]/5. \quad (1)$$

In some regressions we use a measure of the rate of poverty. Since the definition of the rate of poverty differs across countries, we use the uniform measure of poverty rate by the OECD, which limits us to OECD countries only. The data used are more recent, up to 2010, and they include the poverty rate after taxation, namely for disposable income, and the before taxation poverty rate, namely the poverty rate for market income. The two poverty rates are denoted by POVN(j, t) and POVM(j, t), respectively.

For fiscal policy we use a number of measures. The most common measure is public expenditures relative to GDP in percentage terms. We therefore denote by E(j, t) the percentage of expenditures of the public sector in GDP. The moving average over five years of this variable is denoted as in equation (1) by $E_5(j, t)$. The data on fiscal policy in general are taken from the GFS (Government Finance Statistics, total expenditure percent of GDP for total government, supplied from IMF (2014b)). We use mainly cash measures and divide them by GDP (non-cash measures are available for a few countries from 2001

onwards and we use them when cash measures are missing). For the regressions on the rate of poverty, where we restrict the analysis to OECD countries, we use the data on fiscal policy from the OECD database. In addition to total public expenditures we use in our analysis some additional variables of fiscal policy. One such variable is expenditures without defense costs, which we also examine, since defense costs might have a lower effect on redistribution. Since this subtraction of defense costs has no significant effect on the results, we use overall public expenditures thereafter. Public expenditures net of defense are denoted by EMD(j, t). We also test directly the effects of the main tools of redistribution, namely direct taxes, and some of the social transfers. Direct taxes are denoted DT(j, t). Another variable we use in the general regressions is social benefits, SB(j, t), which counts mainly for unemployment insurance payments, all in percentages of GDP.[3] Our analysis of the rate of poverty, using OECD data, enables us to consider a much finer array of social transfers. We use pension spending, denoted P(j, t), spending on income allowances I(j, t), and other labor market interventions L(j, t). All are in percentages of GDP. We also control for the overall social spending in the OECD countries, which includes these transfers plus spending on health and housing.

In addition to the variables that measure inequality and fiscal policy, we use more variables to control for various effects on fiscal policy, where we confront the possibility of reverse causality. One such variable is the state of development of the country. We measure it simply by calculating the ratio of output per capita in country j at time t to output per capita in the US at time t and we denote this ratio D(j, t).[4] The levels of Output per capita for all countries are PPP adjusted and are taken from the WPT data set (Feenstra et al., 2013). Another variable, taken from the same database, which is used to control for openness, is the share of exports and imports in GDP, and it is denoted OPEN(j, t). Other variables used are ethnic fractionalization, as collected by Alesina et al. (2003), and age dependency ratio, namely the ratio of very young and old to the working age population in ages 15–64 (World Bank, 2014). These variables are denoted by FRAC(j, t) and DEP(j, t) respectively.

7.3 The effect of fiscal policy on redistribution: the model

In order to examine how fiscal policy affects redistribution we treat both the distribution of market incomes and the fiscal policy as given and our dependent variable is the distribution of disposable income. This is the result of the interaction of the original distribution of market incomes and of fiscal policy, which changes income through taxes and transfers. Hence, our basic regression model is:

$$GN(j, t) = a + b\ GM(j, t) + c\ E(j, t) + v(j, t), \tag{2}$$

where a is the constant and v(j, t) is a random variable. Note that the distribution of market incomes is assumed to be given in equation (2). This is, of course, only a simplifying assumption and this distribution is affected by many variables, among them even by fiscal policy, through expenditures in education mainly. This is a valid point but it will be taken care of in another research, which we intend to pursue in the near future. The assumption that fiscal policy is also given is dealt with in section 7.5, which examines the possibility of reverse causality.

In the estimation of equation (2) we expect to estimate a positive value of b, actually close to 1, and a negative value of c. The reasons are that inequality of disposable incomes should reflect the basic inequality of market incomes, but it should be reduced by the public sector, mainly by direct taxation and by subsidies. Our main assumption in equation (2) is that we can use the size of the public sector E as a proxy for these two policy variables. Clearly the size of public expenditures should be strongly related to the size of direct taxes. We also assume, that governments that spend more on poverty alleviation and on social benefits, tend to spend more in general.

Note that both inequality and fiscal policy tend to fluctuate over time, mainly due to fluctuations of output and income over the business cycle. As a result we resort to use five years' averages in our estimation. Hence the basic regression model is actually:

$$GN_5(j, t) = a + b\ GM_5(j, t) + c\ E_5(j, t) + u(j, t). \quad (3)$$

Note that u(j, t) is actually equal to $v_5(j, t)$. Another model that we examine in addition to (3) is where the Gini of market incomes affects the distribution of disposable incomes with a lag. It is not clear theoretically why that should be the case, except that it might account for gradual adjustment of the distribution of income. Anyway, our empirical tests show that this is indeed a better choice in most regressions. We therefore present a third model:

$$GN_5(j, t) = a + b\ GM_5(j, t-5) + c\ E_5(j, t) + u(j, t). \quad (4)$$

Equation (4) uses a lag of five years in order to avoid overlap of values in the regression.

Hence, equations (3) and (4) will be our main empirical models of redistribution. We will estimate them and some versions of these equations, mainly using direct taxation and transfers instead of total public expenditures.

7.4 Cross-section regressions

Our first estimations of the model are cross-country regressions for the year 2005. Table 7.1 presents the results of these estimations of the regression models (3) and (4). The first regression in Table 7.1 is just the basic model from

Table 7.1 Gini of disposable incomes and public expenditures across countries, 2005

	Dependent Variable: $GN_5(2005)$				
Independent Variables	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
$GM_5(2005)$	0.816***				
	(0.09)				
$GM_5(2000)$		0.935***	0.961***	0.959***	0.993***
		(0.09)	(0.11)	(0.09)	(0.20)
$E_5(2005)$	–0.406***	–0.423***			
	(0.09)	(0.07)			
$EMD_5(2005)$			–0.439***		
			(0.08)		
$DT_5(2005)$				–.493***	–0.671***
				(0.19)	(0.19)
$SB_5(2005)$					–1.188
					(1.40)
CONST.	12.638***	7.583*	5.902	–2.219	–3.460
	(4.97)	(4.43)	(4.88)	(4.03)	(9.64)
Prob>F	0.0000	0.0000	0.0000	0.000	0.000
R-Squared	0.505	0.648	0.667	0.534	0.526
Number of Countries	83	80	60	102	51

Notes:
1 Significance levels of 1% are denoted by ***, of 5% by **, and of 10% by *.
2 Robust standard errors are in the parenthesis.

equation (3). This regression supports the model very strongly. Inequality of disposable income is strongly affected by inequality of market income, but it is changed significantly by public policy. Each percent of GDP added to public expenditures reduces the Gini coefficient of disposable income by 0.4 percentage points. If we remember that public expenditures in the developed OECD countries vary between 35 and 55 percent of GDP, then this variability itself can explain a variation of 8 points in the Gini coefficient. If we put aside the three less developed countries, Mexico, Chile and Turkey, the Gini coefficient in the OECD countries varies between 24 percent and 38 percent – that is, it varies over a range of 14 percentage points. Fiscal policy accounts, therefore, for the variation of 8 percentage points; namely, it explains more than half of the variability of the Gini coefficient in these countries. This is a very significant effect.

In the second regression in Table 7.1 we run the version of the model in equation (4), where we lag the Gini of market income by five years. The results of this regression are similar to the results of the first regression, except that the R^2 is much higher, namely that this change increases the explanatory power of the regression. As a result we next consider this as our basic specification. The third regression replaces public expenditures with public expenditures minus

defense, which reduces the number of countries from 80 to 60, as data on defense are not available for many countries. Since the results are almost the same as in the second regression, we keep using overall expenditures in the rest of the analysis without subtracting defense costs.

The fourth and fifth regressions in Table 7.1 examine the use the size of public expenditures as a proxy for the main redistribution tools, namely for direct taxation and transfer payments. In these regressions we simply use these variables or similar ones instead of total public expenditures. We find that both variables have a negative effect on inequality, namely that they redistribute income. But the effect of direct taxation is significant, while the effect of social benefits is statistically insignificant. Furthermore, adding social benefits to direct taxation reduces R^2, namely the explanatory power of the regression is reduced. This is unsurprising since the variable we use is a small part of transfers, namely unemployment benefits. We conclude, therefore, that the redistributive effect can be assessed much better by use of progressive direct taxation than by social benefits and social subsidies. Below we find that this result is reversed when we focus on the rate of poverty. Note that the absolute effect of direct taxation has a stronger effect on inequality than total expenditures, but this result is unsurprising. Any marginal unit of public spending is only partially financed by direct taxes, while the rest is financed by indirect taxes. Hence direct taxes fluctuate by less than public spending, so their measured effect on inequality should be higher. Overall, regressions (4) and (5) have lower R^2, so their explanatory power is lower than that of total public spending. From here on, public spending will be our main measure of the redistribution of income from market to disposable incomes. As a result regression model (4) remains our central and basic regression model.

Note, that since the coefficient b is close to 1, we can deduce from equation (4) that the decline from market Gini to disposable Gini is explained mainly by total public expenditures as a share of GDP. This is, therefore, a variable that captures many details of fiscal policy. This result is further strengthened below, where we estimate equation (4) over many periods of time.

7.5 Adding more periods of time

The results of section 7.4 are derived from a cross-section regression of countries in the year 2005. Since our data span a much longer time, we wish to use the full period in order to improve the estimation results. We run both a pooled regression and a panel regression over all data with country's fixed effects. It is important to stress that we expect the pooled regression to give us better results than the panel regression. The reason is that countries usually differ with respect to their preferences toward the redistribution of income. Hence, countries differ in their fiscal policies due to these different preferences. These

differences do not disappear over time. As a result, adding countries' fixed effects to the regression might capture these differences, and as a result it will reduce our ability to identify the effect of fiscal policy in reducing inequality. Note also that both the pooled and the panel regressions run in intervals of five years, as we use averages of five years. Table 7.2 presents the results of these regressions. The first regression is the pooled and the second is the panel regression.

The results of the pooled regressions in Table 7.2 are illuminating. The pooled regression presents similar results to the cross-section estimation of the basic model, except that the coefficients are a bit smaller. Both the Gini of market income and public expenditures have a slightly smaller effect on Gini of disposable income, but the effects are in the same direction and the results are very significant. The high R-squared of this regression shows that public spending is a variable that captures most of the redistribution achieved by the fiscal policy. To see this heuristically, note that the standard deviation of Gini of disposable income is 9.8. The standard deviation of Gini of market income (with a lag) is 8. Hence the lagged Gini of market income accounts for 6.4 of the variation of the Gini of disposable income and leaves a variation of 3.4 unexplained. The standard deviation of public spending is 10.6, so that it explains about 3.2 of the variability of Gini of disposable income, namely it explains almost all the remaining unexplained variability.[5]

Note that although the pooled regression uses more observations, they might not be clearly better than the cross-country regressions in section 7.4. The reason is that the quality of the data on income distribution is improving over time. As a result, the average quality of the data in the regression in the year 2005 is higher than the quality of data in the regressions in Table 7.2, which contains also previous periods with lower quality of data. This point must be taken into account in consideration of the different results.

Table 7.2 Pooled and panel regressions of Gini over public expenditures

Regression	**Pooled**	**Panel**
Dependent Variable	$GN_5(t)$	$GN_5(t)$
$GM_5(t-5)$	0.803***	0.208***
	(0.05)	(0.03)
$E_5(t)$	–0.302***	–0.073***
	(0.04)	(0.03)
Const.	10.542***	28.923***
	(2.80)	(1.85)
Prob > F	0.000	0.000
R-squared	0.610	0.574
No. of observations	306	337

Note that the panel regression yields similar results in direction to the pooled regression, but they are much smaller in size. The reason for that is that much of the differences between fiscal policies in the panel are differences between countries with respect to their social preferences.[6] As a result, these are captured by the panel country fixed effects and they reduce the ability of the regression to identify the effect of fiscal policy on redistribution. For this reason we use for the rest of the analysis pooled regressions instead of panel regressions.

7.6 A test of reverse causality

A famous paper by Meltzer and Richard (1981) claims that the rise of public spending in the twentieth century was driven mainly by public pressure to redistribute income. They conclude that public spending should depend positively on inequality, since higher inequality raises public pressure for redistribution and that increases public spending. Thus, such a mechanism should be taken into consideration in our measurement of the effect of public spending on inequality to avoid endogeneity and reverse causality. In this section we account for this possibility and show that the effect of public policy on inequality is still negative and significant, and the causality clearly goes from fiscal policy to inequality of disposable income. This is done by use of the method of 2SLS, Two Stages Least Squares regression.

In the first stage of estimation, we run a regression of the size of the public sector, namely public expenditure relative to GDP, on a number of exogenous variables. The first one is inequality of market incomes, according to the claim of Meltzer and Richard (1981, 1983). Clearly, the variable that should trigger fiscal policy, according to their model, is inequality of market incomes and not of disposable incomes, which is already the result of fiscal policy. In addition to Gini of market income, we control for more variables that might affect fiscal policy. The first is the level of development of the country. Poor countries are supposed to spend a smaller share of income on public services, as they are considered to be a superior good. An interesting hypothesis on the size of the public sector appears in Rodrik (1998), who claims that openness should increase the public sector, since openness exposes a country to higher risk and governments try to reduce risks. We therefore add openness as another variable to our first-stage estimation. Two population variables are also added. Ethnic fractionalization (FRAC) should affect the preferences of the public toward redistribution. The second variable is demographic, the dependency rate (DEP), which measures the ratio between children and old age population to the working-age population. Since both children and old age are in greater need of redistributing policies, either through education or old-age support, this ratio should also have a positive effect on public expenditures.[7] An additional variable that could have an effect on fiscal policy is a time dummy of

0 before 1995 and 1 from 1995 onward. This variable is supposed to capture the possibility that the collapse of the Soviet Union, after 1990, had a negative effect on public spending in many countries. Whether because this event was viewed as a final victory of the capitalist system, or because it reduced the fear of communism, many view this event as an important trigger to a reduction of the "welfare state" in many countries.

Table 7.3 presents the results of the 2SLS analysis of the effect of public expenditures on Gini of disposable income. The analysis is performed both in pooled regressions and in panel regressions with the between estimator, namely on country means. Each analysis is described by two regressions, the first and the second stage. Regressions (1) and (2) are pooled, and (3) and (4) are between. All variables are five-year averages, except for fractionalization and dependency rate, which are stable over time.

The findings of Table 7.3 are very interesting. The results of the first stage are partly according to expectations, but partly surprising. First of all, the Gini of market incomes affects public spending negatively, which contradicts the theoretical claim of Meltzer and Richard (1981). Namely, the higher inequality

Table 7.3 2SLS regression of Gini over public expenditures

Regression	**(1)**	**(2)**	**(3)**	**(4)**
Dependent Variable	**$E_5(t)$ – 1st stage**	**$GN_5(t)$ – 2nd stage**	**$E_5(t)$ – 1st stage**	**$GN_5(t)$ – 2nd stage**
$GM_5(t-5)$	–0.196***	0.737***	–0.208*	0.856***
	(0.07)	(0.05)	(0.12)	(0.09)
$D_5(t)$	4.767***		4.970***	
	(0.70)		(1.16)	
$OPEN_5(t)$	–25.325***		–27.904***	
	(3.89)		(6.28)	
FRAC	–9.817***		–5.823	
	(2.54)		(4.12)	
DEP	0.166***		0.134*	
	(0.06)		(0.09)	
D1995	–0.245			
	(1.04)			
2SLS $E_5(t)$		–.603***		–0.556***
		(0.07)		(0.12)
Const.	37.335***	22.475***	37.809***	15.603***
	(3.55)	(3.41)	(5.73)	(5.52)
Prob>F	0.000		0.000	
Prob>χ^2		0.000		0.000
R-squared	0.315	0.496		
R-squared between			0.352	0.561
No. of observations	304	304	334	334
No. of groups			90	90

of market incomes is correlated with less and not with more redistributive policies.[8] Second, openness affects public spending negatively, and that contradicts the findings of Rodrik (1998). The other variables in the first-stage regression affect public expenditures as expected. The level of development increases the share of public expenditures in GDP, reaffirming that public services can be viewed as a luxury good. The two population variables, ethnic fractionalization and the age dependency ratio, also affect public spending as expected. Ethnic fractionalization reduces redistribution policies, since people in the country feel less solidarity, while a larger share of old and young increases social expenditures. The time dummy variable of 1995 does not have a significant effect on public spending.

The second-stage regressions strongly reaffirm our previous results. The effect of the market Gini on disposable Gini in both pooled and between regressions is close to the effect in the pooled regression in Table 7.2. Most importantly, the implied public spending, which is derived from the first-stage regression, reduces inequality significantly in both regressions (2) and (4). Its effect is larger in size than in Tables 7.1 and 7.2, but since this is only the effect of the explained part of the variable E, it should be double in size, as it is indeed.

7.7 Redistributive fiscal policy and the rate of poverty

In this section we examine the effect of redistributive fiscal policies not on general inequality, but rather on the lower domain of the distribution, namely on the rate of poverty. This is the share in population which is defined as poor in percents. It is hard to get comparable data on poverty rates across all countries, because they differ significantly in terms of their definition of poverty. In this subsection we focus therefore on the OECD countries, since the organization uses a uniform definition of poverty in their statistics. The rate of poverty is defined as the share of people with household income that is lower than half of the median income. The OECD countries differ significantly in their rates of poverty, from close to 5 percent in the Czech Republic and in Denmark, to more than 20 percent in Mexico and in Israel. We use this uniform data on poverty in a cross-section test in 2010 and also in pooled regressions that use also previous periods.

The dependent variable in all regressions is the rate of poverty of disposable incomes. The explanatory variables are the rate of poverty of market incomes and various variables that reflect fiscal policy. To keep consistency of data we use fiscal policy data from OECD and not from GFS, as done in previous sections. The first fiscal variable is the total public expenditures as percent of GDP, E(j, t). A second variable is the sum of public social expenditures, which comprises of pensions, income support subsidies, unemployment benefits, health expenditures by the public sector and housing. This variable is denoted SOC(j, t) for country j in year t. A third variable is public pensions, denoted

PEN(j, t), which should affect poverty significantly since they increase income of the old, which are usually much poorer than the rest of the population. A fourth variable is income support to working and non-working people. This variable is denoted SUP(j, t). All variables are given in terms of percentage of GDP. We first present the results of the cross-section regression in Table 7.4 and then of the pooled regressions. All data for the regressions in this section are taken from the database of OECD.

Table 7.4 presents the results of five cross-country regressions. The number of countries is around 30, but it differs from one regression to the other due to data limitations. Regression (1) is the basic regression of the rate of poverty in disposable incomes on the rate of poverty in market income and on total public expenditures. There are two interesting results in this regression. The first is that the relation between the rate of poverty of disposable income and of market income is weak and hardly significant. This stands in contrast with the results in sections 7.4 and 7.5 that show that the relation between Gini disposable and Gini market is very significant and much larger, close to 1. The reason is that the redistribution of income by fiscal policy principally affects people with low incomes and succeeds in reducing the rate of poverty significantly. This is done mainly by the social transfers that target mainly poor populations. Regression (1) in Table 7.4 also shows that public expenditures reduce

Table 7.4 Cross-section tests of the rate of poverty on fiscal policy in OECD countries

	Dependent Variable: $POVN_5$(2010)				
Independent Variables	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
$POVM_5$(2010)	0.226*	0.218*			0.208*
	(0.14)	(0.14)			(0.13)
E_5(2010)	–0.316***		–0.081	–0.081	
	(0.12)		(0.15)	(0.15)	
SOC_5(2010)		–0.495***			
		(0.14)			
PEN_5(2010)				0.005	–0.257
				(0.237)	(0.20)
SUP_5(2010)			–1.365***	–1.361***	–1.692***
			(0.43)	(0.47)	(0.36)
CONST.	18.301***	14.682***	19.691***	19.718***	13.948***
	(5.26)	(3.48)	(4.35)	(4.62)	(3.11)
Prob>F	0.0455	0.0053	0.0006	0.0023	0.0008
R-Squared	0.227	0.312	0.433	0.433	0.456
Number of Countries	27	31	29	29	31

Notes:
1 Significance levels of 1% are denoted by ***, of 5% by **, and of 10% by *.
2 Robust standard errors are in the parenthesis.

significantly the rate of poverty. Every additional percent of GDP that goes to the public sector reduces the rate of poverty by 0.32 percent. The next regressions examine the effects of partial public expenditures that are more directed to poverty alleviation.

Regression (2) replaces the total public expenditures by social expenditures (SOC), which is defined above. Note that in the OECD countries this is a large part of public expenditures (40 percent to 50 percent of total expenditures), but it is clearly more focused on the poorer parts of the population. We should therefore expect it to have a much stronger effect on the rate of poverty. Indeed, this regression supports this hypothesis. Every percent of GDP that goes to social public expenditures reduces the rate of poverty by half a percentage point. Regression (3) examines the effect of a more specific expenditures and that is income support subsidies. It shows that once this expenditure is added, total expenditures become insignificant and the effect of income support is much stronger. Every percent of GDP that goes to income support subsidies reduces the rate of poverty by 1.4 percentage points – more than three times the effect of all public expenditures. In regression (4), we add the variable of expenditures on pensions and it does not change the results of regression (3) at all. Pensions themselves have no effect on the rate of poverty, unlike our preliminary conjecture. Their effect is extremely small, positive and it is not significantly different from 0.

An interesting result that comes up from Table 7.4 is that total public expenditures become ineffective once income support is controlled for in the regression, as in regressions (3) and (4). As a result, regression (5) omits the total public expenditures and adds to the regression instead the market income rate of poverty, which has a relatively weak effect on the disposable income rate of poverty. Clearly, regression (5) seems to be the best regression in Table 7.4. It has the higher value of R-squared and its estimates of the coefficients have the higher significance. According to regression (5), the effect of income support on poverty reduction is the strongest. Every percent of GDP increase in income support reduces the rate of poverty by 1.7 percentage points.

Interestingly, although the marginal effect of total public expenditures on the poverty rate is insignificant, as demonstrated by regressions (3) and (4), its total effect is large. This can be seen in Table 7.4 by the size of the constant. When the variable E (or SOC) is not included in the regression the size of the constant is around 14 percent, but when E is included, in regressions (3) and (4) the size of the constant increases to close to 20 percent. This means that the overall effect of total public expenditures on reducing the rate of poverty is around five percentage points. This is a sizable effect, even if it does not come up significant in the regression.

Table 7.5 presents similar tests to those detailed in Table 7.4, but conducted for a larger data span, that goes back in time to the 1990s. Of course, not all

Table 7.5 Pooled regressions of the rate of poverty on fiscal policy in OECD countries

	Dependent Variable: $POVN_5$(j, t)				
Independent Variables	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
$POVM_5$(j, t)	0.197***				0.335***
	(0.08)				(0.09)
$POVM_5$(j, t–5)		0.091	0.290***	–0.265**	
		(0.10)	(0.10)	(0.13)	
E_5(j, t)	–0.375***	–0.426***			
	(0.05)	(0.06)			
SOC_5(j, t)			–0.599***		–0.988***
			(0.08)		(0.16)
PEN_5(j, t)					0.545***
					(0.19)
L_5(j, t)				–2.247***	0.507
				(0.30)	(0.41)
CONST.	21.569***	27.101***	15.604***	21.340***	16.380***
	(2.57)	(4.41)	(2.78)	(4.62)	(2.59)
Prob>F	0.0000	0.0000	0.0000	0.0000	0.0000
R-Squared	0.468	0.495	0.463	0.563	0.650
Number of observations	77	56	75	47	74

Notes:
1 Significance levels of 1% are denoted by ***, of 5% by **, and of 10% by *.
2 Robust standard errors are in the parenthesis.

countries have full data going back in time as others and the data set is not balanced. This does not bother us too much, as we abstract from any country fixed effect and focus only on the observed differences in fiscal policy. A major difference between the two tables is that in Table 7.5 we no longer use the income support, due to missing data, but instead rely on a variable that includes all labor market support, which is denoted by L(j, t).

The pooled regressions in Table 7.5 strengthen the conclusions we derive from the results of the cross-section regressions in Table 7.4. The effect of total public spending on poverty reduction is larger and every percent of GDP in public spending reduces the rate of poverty by 0.4 percentage points. This is a strong effect. As we narrow our focus on expenditures directed to the poor their effect becomes stronger. Every percent of GDP in social spending reduces the rate of poverty by 0.6 percentage points. The stronger effect is observed with respect to labor market subsidies. Every percent of GDP in these subsidies reduces the rate of poverty by 2.25 percentage points. This is indeed a very strong effect. Note that in regression (5), which includes all social fiscal variables, some of them "swallow" the effect of others. Hence, all poverty reduction is reserved to total social expenditures, which affect poverty by much more than in regression (3), where social expenditures come alone.

7.8 Summary and conclusions

In this paper we examine what determines inequality in a society, and focus mainly on fiscal policy and its redistribution of income from market income to disposable income through direct taxation and various subsidies. Our study is purely empirical and it uses international comparisons. Hence it consists of cross-country regressions or of pooled regressions and panel regressions, when more periods of time are added.

Our empirical analysis points very clearly at fiscal policy as a central determinant of inequality. It plays a major role in the redistribution of income from market incomes to disposable incomes, through direct taxation and through social subsidies, which are both highly progressive. We measure this effect of fiscal policy and find out that it is quite large. Every increase of fiscal spending of the size of one percent of GDP reduces the Gini coefficient by 0.4 percentage points. Since inequality of market incomes is quite similar across countries, the large differences we observe in inequality are due mainly to differences in fiscal policy across countries. We also find that fiscal policy has a strong effect on poverty reduction, from market incomes to disposable incomes. Here income support subsidies and similar interventions in the labor market, which are focused on the poor, play an especially large role.

Finally, our study can also hint at one possible explanation for the rise of inequality in recent decades. In addition to the standard explanations, such as globalization and skill-biased technical change, our paper suggests an alternative mechanism, namely a reduction in the size of the public sector. According to our study, this can increase inequality significantly. This final observation leads us to another question – namely, can fiscal policy and similar public policies, like labor market interventions, be part of the overall determinant of inequality, not only of disposable income, but of market incomes as well? We plan to devote our next research to this issue.

Notes

1. The regression coefficient of the Gini over public expenditures is –.35 and is significant at 1%. It means that if public expenditures rise by 1 percent of GDP, Gini of disposable incomes declines by 0.35 points.
2. Muinelo-Gallo and Roca-Sagalés (2013) also treat fiscal policy as endogenous, but they do it in a system of three equations, of growth, inequality and fiscal policy, and they do not use our two-stage approach.
3. Data on social benefits, taxes and defense as shares of GDP are all from IMF (2014b).
4. We also tried an alternative measure for development, the share of agriculture in GDP, which yields similar results. We do not use it since it is available for fewer countries.

5. We also examined the use of the lagged value of the Gini of market income in the pooled regression. Using instead the contemporaneous Gini of market income the results are very similar to Table 7.2. The coefficient of contemporaneous Gini is 0.839, the coefficient of public spending is –0.35 and the R-squared is 0.57.
6. This is also shown below in the first-stage regression in section 7.6.
7. Interestingly, Muinelo-Gallo and Roca-Sagalés (2013) use very similar variables in their estimation of fiscal policy. They do not use fractionalization, but use political system data instead. Another difference is that we use only aggregate fiscal expenditures as the dependent variable, while they use various parts of fiscal policy instead. The main difference between our estimation and theirs is that we use the results of this regression for testing reverse causality, which is not done by Muinelo-Gallo and Roca-Sagalés (2013).
8. Interestingly, this result is reached also in Muinelo-Gallo and Roca-Sagalés (2013).

References

Alesina, Alberto, Devleeschauwer, Arnaud, Easterly, William, Kurlaut, Sergio and Wacziarg, Romain (2003) "Fractionalization," *Journal of Economic Growth*, vol. 8, pp. 155–94.

Atkinson, Anthony B. (1997) "Bringing Income Distribution in From the Cold," *The Economic Journal*, vol. 107, pp. 297–321.

Barro, Robert J. (2000) "Inequality and Growth in a Panel of Countries," *Journal of Economic Growth*, vol. 5, pp. 5–32.

Chu, Ke-young, Davoodi, Hamid R., and Gupta, Sanjeev (2004) "Income Distribution and Tax and Government Social-Spending Policies in Developing Countries," in Giovanni A. Cornia and Julius Court (eds), *Inequality, Growth, and Poverty in an Era of Liberalization and Globalization*. Oxford: Oxford University Press.

Feenstra, Robert C., Inklaar, Robert, and Timmer, Marcel P. (2013) "The Next Generation of the Penn World Table." Downloadable at www.ggdc.net/pwt.

International Monetary Fund (IMF) (2014a) *Fiscal Policy and Income Inequality*, IMF Policy Paper, January.

International Monetary Fund (IMF) (2014b) *Government Finance Statistics*, Database.

Kuznets, Simon (1955) "Economic Growth and Income Inequality," *American Economic Review*, vol. 45, pp. 1–28.

Martinez-Vazquez, Jorge, Vulovic, Violeta, and Moreno-Dodson, Blanca (2012) "The Impact of Tax and Expenditure Policies on Income Distribution: Evidence from a Large Panel of Countries," *Hacienda Pública Española*, vol. 200, pp. 95–130.

Meltzer, Allan H., and Richard, Scott F. (1981) "A Rational Theory of the Size of Government," *Journal of Political Economy*, vol. 89, pp. 914–27.

Meltzer, Allan H., and Richard, Scott F. (1983) "Tests of a Rational Theory of the Size of Government," *Public Choice*, vol. 41, pp. 403–18.

Muinelo-Gallo, Leonel, and Roca-Sagalés, Oriol (2013) "Joint Determinants of Fiscal Policy, Income Inequality and Economic Growth," *Economic Modeling*, 30, 814–24.

Niehues, Judith, (2010) "Social Spending Generosity and Income Inequality: A Dynamic Panel Approach," IZA Working Paper no. 5178.

OECD (2014) "Employment Protection Annual Time Series Data 1985–2013." Database.

O'Mahony, Mary, and Timmer, Marcel P. (2009) "Output, Input and Productivity Measures at the Industry Level: the EU KLEMS Database", *Economic Journal*, 119 (538), F374–F403.

Ospina, Monica (2010) "The Effect of Social Spending on Income Inequality: An Analysis for Latin American Countries," Center for Research in Economics and Finance (CIEF) Working Papers, No. 10-03.

Rodrik, Dani (1998) "Why Do More Open Economies Have Bigger Governments?," *Journal of Political Economy*, 106, 997–1032.

Solt, Frederick (2009) "Standardizing the World Income Inequality Database," *Social Science Quarterly*, 90, 231–42.

UNU-WIDER (2008) *World Income Inequality Database*, Version 2.0c, May 2008. Available at hhttp://www.wider.unu.edu/research/Database/en_GB/database/i.

Visser, Jelle (2013) "Database on Institutional Characteristics of Trade Unions, Wage Setting, State Intervention and Social Pacts, 1960–2011," ICTWSS.

World Bank (2014) *World Development Indicators*. http://data.worldbank.org/data-catalog/world-development-indicators/ Accessed on 10 October 2014.

Comments on "The Effects of Fiscal Redistribution" by Michele Battisti and Joseph Zeira

James E. Foster
George Washington University

One reason that disposable income is preferred over market income in distributional evaluations is that it takes into account taxes that lower the resources and transfers that raise them, and hence is a better estimate of the actual resources available to people. To be sure, there are many other relevant factors that disposable income does not include. For example, it does not account for the direct provision of goods and services by governments, including in-kind transfers, which can free up monetary resources for other uses while enhancing people's capabilities. Savings and wealth are not being considered, which significantly alter a person's vulnerability to shocks and ability to muster monetary resources if needed. It ignores the non-monetary resources and capabilities a person has, which directly impact a person's wellbeing while affecting the ability to make use of other resources, including monetary resources.[1] Indirect taxes of various types are also typically not deducted in deriving disposable income. Nonetheless, disposable income is the natural place to begin the journey from income to welfare, and even more, a comparison of disposable and market income distribution can help clarify how government and market institutions are shaping inequality (and poverty) across countries.[2]

Battisti and Zeira study the disposable (or "post-fiscal") income distribution for a range of 80 economies, including the OECD countries.[3] They motivate their study using a graph that plots each country's 2005 level of inequality, as measured by the Gini coefficient of the distribution of disposable income (the "post-fiscal Gini"), against the size of government relative to the economy in the same year, and notice a distinctly negative relationship. In other words, countries having a larger government share in GDP also tend to have lower inequality in disposable incomes. This raises a number of natural questions: Is this negative relationship maintained when other factors are included? Does the relationship hold over time as well? How much of this is due to market incomes versus government policies?

To address these questions, they regress the post-fiscal Gini on the market Gini and the government share in GDP. A strong negative relationship is found

between the government share and the post-fiscal Gini (namely, that every additional percent of government in GDP reduces the post-fiscal Gini by one percent).[4] They then turn to the question of reverse causality and reject the converse hypothesis (often found in discussions of political economy) that higher inequality necessarily leads to a greater government share. They further examine the robustness of their main result by including data on additional years, first pooling the observations and then constructing a panel of countries through time. The pooled results are similar to the cross-section results, but while the panel regression suggests that post-fiscal inequality is still negatively related to the relative size of government, the coefficient is much smaller. Inequality in disposable income is much less responsive to changes in the relative size of government over time than was found in the cross section. Instead, a large part of the cross-country variation is now attributed to country fixed effects.

They conduct similar exercises for poverty measures derived from the post-fiscal and market income distributions. However, since the data are now limited to 30 OECD countries, the definition used is a relative poverty measure (namely, the headcount ratio at the relative poverty line of one-half the median income), and hence the poverty is likely closer to relative inequality than traditional notions of absolute poverty. The authors regress the post-fiscal headcount ratio on the market headcount ratio and the government share in GDP for their now smaller data set covering 30 countries. The results indicate that a larger government share reduces the poverty rate in such a way that each additional percent of GDP lowers the poverty rate by about one-third of a percentage point. The authors note that the strength of association between post-fiscal and monetary poverty is much smaller than was observed for inequality.

The size of government variable can be broken down further into different forms of taxation or different categories of expenditures in order to identify which components of fiscal policies seem to be driving the conclusions. This they do, and although the data used in the analyses of inequality and poverty are different, a coherent picture seems to emerge. The change in inequality from market to post-fiscal inequality seems to be best explained by direct taxation, while the change in poverty is most strongly linked to income subsidies. The government clearly plays a large role in moderating the levels of inequality and poverty delivered up by the market economy.

This is focused, well-written paper that has a clear message concerning the special role of government in altering market outcomes in the distribution of income. Of course, a complete analysis might include the expenditure side of the equation (e.g., the value of publicly provided goods and services, including in-kind transfers), or account for the various assets a person has control over, or perhaps even account for the capabilities of people and how they are affected by government policies. But then this would depart from the main objective of this paper, which is to provide some insight into why some countries appear to

be more successful in reducing inequality and poverty than others. While I am not fully convinced that "reduction in the size of the public sector" is in itself a "possible explanation to the rise of inequality in recent decades," I would agree with the authors that their results hint strongly in this general direction.

There are several items in the paper that could benefit from some additional discussion. First, as noted in the paper, the present work is rather closely linked to a longstanding literature on the effective progressivity (or redistribution) of fiscal policy. The use of measures based on the Gini coefficient dates back at least to Musgrave and Thin (1948), while the intuitive measure frequently used – the market Gini minus the post-fiscal Gini – is attributed to Reynolds and Smolensky (1977). There are many analyses that use this index to measure redistribution, the most recent of which is IMF (2014) as cited by the authors. It would be quite useful to better understand the benefits and costs of regressing the post-fiscal Gini on the market Gini (as the authors do) instead of examining their difference (as is more commonly done). The latter approach is indeed intuitive and as noted in IMF (2014) there is a significant body of analysis available – much of which parallels and augments the key results of this paper. The authors emphasize that using the difference in Ginis in regression equations (say, by regressing the Reynolds–Smolinski measure on various fiscal policies) will effectively restrict the coefficient on the market Gini to be 1. While this may be true, it is not entirely clear why the restriction represents a significant cost, given the added intuition from using a measure of fiscal redistribution and, indeed, given the authors own empirical findings of a coefficient quite close to 1.[5]

Second, most modern distributional analyses employ multiple inequality or poverty measures to reduce the possibility that the results depend purely on the particular measure that has been chosen. The Gini coefficient is the most commonly used index, but there are several good reasons for not using it, not the least of which is its wholesale violation of subgroup consistency (see Cowell, 1988, or Foster and Sen, 1997). Likewise, the poverty measure being used here – the headcount ratio – is a remarkably crude way of evaluating gains and losses in poverty. Add to this the fact that a relative poverty line of half the median income can lead to some very unintuitive comparisons (where a drop in measured poverty is consistent with a everyone's becoming poorer), and we are left wondering what would happen if we used an absolute poverty line or a poverty measure that is sensitive to depth or severity (e.g., Foster, Greer, Thorbecke, 1984). Of course, data constraints may well prevent the kinds of robustness analyses suggested here, but it is important to bear in mind, and take note of, this limitation.

Key components of wellbeing are not represented in income statistics and hence conclusions about distributional issues, such as poverty, inequality, or the inclusiveness of growth, may require other dimensions to be considered at

the same time. For example, the Multidimensional Poverty Index (MPI) published annually by the United Nations for over 100 countries, includes information on the dimensions of education and health, as well as standard of living. It can identify very different households as poor when compared to standard monetary approaches (such as the $1.25 figures produced by the World Bank) and consequently provides an alternative global picture of poverty.[6] Since many government policies are designed to directly impact non-monetary dimensions of wellbeing, multidimensional measures that reflect these dimensions reveal the actual situations of people, and the impact of policies through non-income dimensions. It would be natural to expand the incidence analysis to include multidimensional measures.

Third, Lustig and Higgins (2012) criticize the standard measures of inequality and poverty for being entirely anonymous, and hence unable to track the identities of winners and losers.[7] They propose using an analogue of mobility matrix to evaluate fiscal incidence, and note the great heterogeneity of incidence among persons who are initially quite similar. For example, they note that 15 percent of the poor in Brazil (as reckoned using market income) become classified as extremely poor in disposable income as a result of the overall fiscal system – which includes regressive indirect taxes. Their analysis also points to an alternative approach – country analyses rather than cross-country regressions that require major assumptions in order to be applied.

Fourth, the Solt (2009) data set being used here is of a rather different nature than other sources, since its summary statistics on income distribution are generated by an imputation model having many assumptions, rather than being estimated directly from data. This, naturally, has led to a number of critiques, many of which are summarized in Jenkins (2014). He notes for example that the imputation method employs a smoothing algorithm, yielding results that contradict figures derived directly from existing high quality datasets; he also suggests that regression results obtained with the data could well be biased in unpredictable ways. These cautionary descriptions should be borne in mind.[8]

A final observation focuses on the structure of inequality measures to suggest an alternative indicator for evaluating fiscal policy. It turns out that nearly every inequality measure in common use is composed of two "income standards" each of which summarizes the entire distribution using a representative income and satisfies a number of basic axioms.[9] One of the two income standards focuses on lower incomes and therefore takes values that are below the second, higher income standard. Inequality is measured using an expression that compares the first income standard to the second. Measured inequality can be reduced by raising the lower income standard or by lowering the higher income standard, thus lowering the relative distance between the two.

In the case of the Gini coefficient, the lower income standard is the Sen welfare measure, S, a distribution-sensitive income standard that lies below the

mean.[10] The second income standard is the mean μ itself. The Gini coefficient can be expressed as $G=(\mu-S)/\mu=1-S/\mu$. Given that the post-fiscal Gini G^* is lower than the market Gini G, this indicates that S^*/μ^* exceeds S/μ, where each ratio is an index of equality. If the mean impact on incomes is 0, this ensures that the post-fiscal welfare as measured using the Sen welfare measure is greater than the welfare of the market income distribution. However, if overall taxes exceed transfers, so that $\mu^*<\mu$, then inequality can be falling without a concomitant increase in welfare. One approach would be to ignore the mean μ entirely and use the low-income standard S in evaluating the combined impact of taxes and transfers on people's welfare. Indeed, other distribution sensitive lower income standards, such as Atkinson's equally distributed equivalent income and the World Bank's mean of the lower 40%, could perform the same function of ensuring that the fiscal policies are welfare improving. Naturally, data availability may restrict the choice of income standard; it can be noted that S can be constructed whenever there is information on the Gini coefficient and the mean.

Notes

1. For a discussion of the capability approach, see for example Sen (1992), Foster and Sen (1997) or Basu and Lòpez-Calva (2011).
2. These issues fall within the area of fiscal incidence analysis in public finance, studies the combined effects of taxes and transfers on people's income and welfare; see for example Reynolds and Smolensky (1977).
3. The number of countries varies for different analyses in the paper, due to data constraints. It would be helpful if the authors were to post the lists of countries and dates used in each regression.
4. In one direction this qualitative result is rather obvious – since a very small public sector would likewise limit the real possibilities for redistribution. However, while necessary, a larger share of government in GDP would not be sufficient. The impact depends on *who* is paying the taxes and *who* is receiving the income transfers, and hence on the redistributional qualities of the fiscal system.
5. There is also the problem raised by Foster (2015) that the values of inequality measures may not, in fact, be cardinally meaningful, which would call into question their use in regression equations, and also in Reynolds–Smolensky type indices of redistribution. Poverty measures like the headcount ratio avoid this drawback.
6. See Alkire and Santos (2013) for a discussion of the MPI.
7. See the approach of Kakwani (1977) who uses the non-anonymous concentration coefficient over disposable incomes instead of the anonymous Gini coefficient.
8. The smoothing algorithm underlying the Solt (2009) data set leads one to question the necessity or meaning of the five-year averaging employed by Battisti and Zeira in equation (3).
9. See Foster et al (2013).
10. See the definition of S in Foster and Sen (1997).

References

Alkire, S., and Foster, J.E. (2011) "Counting and Multidimensional Poverty Measurement" (with S. Alkire), *Journal of Public Economics*, vol. 95, iss. 7–8, pp. 476–87.

Alkire, S., and Santos, M.E. (2013) "A Multidimensional Approach: Poverty Measurement & Beyond," *Social Indicators Research*, vol. 112, pp. 239–57.

Basu, K., and Lòpez-Calva, L.F. (2011) "Functionings and Capabilities," in K.J. Arrow, A. K. Sen, and K. Suzumura (eds), *Handbook of Social Choice and Welfare*, Volume 2. Amsterdam: Elsevier, pp. 153–87.

Cowell, F. (1988) "Inequality Decomposition: Three Bad Measures," *Bulletin of Economic Research*, vol. 40, iss. 4, pp. 309–12.

Foster, J.E. (2015) "A Framework for Measuring Inclusive Growth," mimeo, George Washington University.

Foster, J.E., Greer, J., and Thorbecke, E. (1984) "A Class of Decomposable Poverty Measures," *Econometrica*, vol. 52 (May), pp. 761–6.

Foster, J.E. and A.K. Sen (1997) "On Economic Inequality: After a Quarter Century," Annex to the enlarged edition of *On Economic Inequality* by Amartya Sen, Oxford: Clarendon Press.

Foster, J., Seth, S., Lokshin, M., and Sajaia, Z. (2013) *A Unified Approach to Measuring Poverty and Inequality: Theory and Practice*. Washington, DC: World Bank Press.

International Monetary Fund (IMF) (2014) *Fiscal Policy and Income Inequality*, IMF Policy Paper, January.

Jenkins, S.P. (2014) "World Income Inequality Databases: An Assessment of WIID and SWIID," IZA Discussion Paper No. 8501.

Kakwani, N.C. (1977) "Measurement of Tax Progressivity: An International Comparison," *The Economic Journal*, vol. 87 (March), pp. 71–80.

Lustig, N., and Higgins, S. (2012) "Fiscal Incidence, Fiscal Mobility and the Poor: A New Approach," Tulane Economics Working Paper Series, Number 1202.

Musgrave, R.A. and Thin, T. (1948) "Income Tax Progression, 1929–48," *Journal of Political Economy*, vol. 56, no. 6, pp. 498–514.

Reynolds, M., and Smolensky, E. (1977) *Public Expenditures, Taxes, and the Distribution of Income: The United States, 1950, 1961, 1970*. New York: Academic Press.

Sen, A. (1992) *Inequality Reexamined*. Oxford: Clarendon Press.

Solt, F. (2009) "Standardizing the World Income Inequality Database," *Social Science Quarterly*, vol. 90, pp. 231–42.

8

Inequality of Happiness: Evidence of the Compression of the Subjective-Wellbeing Distribution with Economic Growth

John Ifcher
Department of Economics, Santa Clara University

Homa Zarghamee
Department of Economics, Barnard College, Columbia University

8.1 Introduction

The use of Subjective Wellbeing (SWB) measures in economics research has grown markedly (Kahneman and Krueger 2006). This has come about for at least two reasons. First, the measures have been systematically validated as reliable for examining a range of questions. Second, economists have long relied on income as a proxy for wellbeing. However, research shows that there are potentially large slippages between economic indicators and wellbeing (Diener and Seligman 2004). Thus, SWB measures have become an important alternative proxy for wellbeing. Indeed, SWB measures have also caught the attention of policy makers. The OECD launched the Better Life Index in 2011 as an alternative wellbeing measure; and the former French President Nicolas Sarkozy formed the Stiglitz Commission in 2008 to identify the limits of gross domestic product (GDP) as a measure of wellbeing and to identify alternative measures (Stiglitz, Sen, and Fitoussi 2010).

When studying the distribution of income, economists have long recognized the importance of examining measures of central tendency and dispersion, as the latter are necessary to understand income inequality and poverty (Stiglitz, Sen, and Fitoussi 2010). Thus, there is a vast literature analyzing both the first and second moments of the distribution of income. For example, the Lorenz and Kuznets curves try to model the distribution of income, and the Gini coefficient summarizes the entire distribution in a scalar (see Atkinson 1970; Gastwirth 1972; Gini 1921; Gottschalk and Smeeding 1997; Kuznets 1955; and Lorenz 1905). In contrast, the vast majority of SWB research focuses on mean SWB. Given the current interest in SWB measures, and recognizing that

the entire distribution of SWB merits study, we believe it is important to study SWB inequality (dispersion) as well as mean SWB.

In this paper, we contribute to the emerging SWB literature by investigating the relationship between economic growth and SWB inequality using data from the World Values Survey (WVS) and the World Bank's World Development Indicators (WDI). The results suggest that economic growth is inversely related to SWB inequality in cross-sectional analysis. There is also some evidence from time series analysis that countries that experience greater economic growth rates also experience the greater decreases in SWB inequality, although this pattern does not hold for two of the fastest-growing countries in the data set. This is important because it indicates that economic growth may reduce SWB inequality over time, even if it does not increase mean SWB. The paper proceeds as follows. Section II reviews the related literature. Section III describes the data. Section IV presents the results. Section V concludes.

8.2 Literature review

The vastness of the income inequality literature illustrates the importance of studying income's distribution (for example, Atkinson 1970; Gastwirth 1972; Gini 1921; Gottschalk and Smeeding 1997; Lorenz 1905). In contrast, there are only a few papers that have studied SWB inequality.[1] Stevenson and Wolfers (2008b) examine trends in happiness inequality in the United States from 1972 to 2006 using the General Social Survey. They find that happiness inequality decreased during this period. The authors juxtapose their finding with the concurrent rise in income inequality in the United States but do not examine the relationship between happiness inequality and economic growth.

Easterlin (2012) studies SWB inequality in developed capitalist countries and in countries that transitioned from socialism to capitalism using the WVS. He finds that developed capitalist countries (with the exception of the Nordic welfare states) had greater SWB inequality than "Soviet-style" socialist countries before the transition. This pattern reverses after the transition, with the increase in SWB inequality in former socialist countries resulting from decreased SWB among low-income individuals. Easterlin et al. (2012) find the same pattern in China after the restructuring of state-owned enterprises (SOEs) and trimming of social safety nets.

Finally, Veenhoven (2005a) attempts to refute the "The Great U-Turn:" the return of social inequality in modern society.[2] Veenhoven examines trends in SWB inequality using the standard deviation of life satisfaction between 1973 and 2001 in Eurobarometer data. He shows that in that time period SWB inequality decreased. In his analysis, Veenhoven does not examine the relationship between SWB inequality and economic growth. He does, however, examine the relationships between SWB inequality and modernity using data

from the WVS. To do so, he plots SWB inequality against several measures of modernity, such as purchasing power, freedom in private life, urbanization, and education. He concludes that as countries "modernize," SWB inequality decreases. His analysis, however, is limited to a cross-sectional analysis using only two waves of data from the WVS. Our paper builds upon Veenhoven's (2005a) paper in two important ways. First, we use all five waves of the WVS, and second, we compare SWB inequality and per capita GDP (GPDpc) using both cross-sectional and time series analysis. This is important since, as discussed below, many researchers believe that the relationship between mean SWB and GDPpc is different in cross-section than in time series.

Cross-sectional analysis indicates that there is a positive relationship between mean SWB and GDPpc within a country and also across countries (Easterlin 1974; Stevenson and Wolfers 2008a; and Stevenson and Wolfers 2013). That is, within a country, individuals with higher income have higher SWB, on average, than individuals with lower income; and countries with higher average income have higher mean SWB. However, many researchers believe that this relationship vanishes in time series; this is the "Easterlin Paradox," introduced in Easterlin (1974) (see also Easterlin 1995; Easterlin 2013; and Easterlin et al. 2010). Various explanations have been proposed for the divergent results, for example, that relative income, not absolute income, is associated with SWB, or that individuals adapt to higher income over time. After the publication of the Easterlin Paradox, a heated debate has developed regarding the validity of Easterlin's finding, as many find it hard to believe that mean SWB does not increase with per capita income within a country over time. To determine if the paradox exists, the important variable to consider appears to be the time frame of the analysis. When Easterlin first proposed the paradox, he found that in long-term time series, the correlation between mean happiness and per capita income disappeared. A thorough critique of the paradox is by Stevenson and Wolfers (2008a), who examine multiple shorter time series to demonstrate that the association between mean happiness and per capita income does exist. The main difference between these two analyses is that Stevenson and Wolfers consider shorter time series and Easterlin considers longer time series. It is important to note that the existence of the Easterlin Paradox is a subject of active debate (for example, Stevenson and Wolfers 2008a; Stevenson and Wolfers 2013; and Easterlin et al. 2010). For the purpose of this paper, we can remain agnostic.

We contribute to the SWB inequality literature by performing a systematic analysis of the relationship between SWB inequality and economic growth. We examine the relationship between SWB inequality and economic growth in both a cross-sectional and time series analysis. Because our results are for the most part consistent across these two analyses, they do not present the challenge that the Easterlin paradox does. Our research also suggests that, despite

the controversy the Easterlin paradox presents, there may be an additional benefit from increasing per capita income within a country: namely, decreasing SWB inequality.

8.3 Data

The SWB data for this study come from the WVS, the most comprehensive data set, in terms of years and countries covered, available for studying SWB. It has been administered five times. The first wave, administered between 1981 and 1984, includes 21 countries and the fifth wave, administered between 2005 and 2008, includes 56 countries. In total, there are over 350,000 respondents; the survey has been administered in 98 countries at least once; and there are 248 country-wave pairs (for example, the United States – Wave 1).

The WVS includes a standard Life-Satisfaction (LS) question as well as a happiness question. The former asks: "All things considered, how satisfied are you with your life as a whole these days?" where "1" is defined as "dissatisfied" and "10" is defined as "satisfied;" Figure 8.1a presents a histogram of the LS data. Like Veenhoven's study, our analysis uses the standard deviation of the LS question rather than the happiness question. Its response scale is larger (10 possible responses versus 4) and the standard deviation is greater than is the standard deviation of happiness (2.45 versus 0.74). Further, the LS question is believed to be better for making cross-country comparisons than the happiness question (Di Tella et al. 2010). Of the 248 country-wave pairs in the WVS, there is LS data for 246; LS data are missing for Korea 1996 and Pakistan 1997. Also, Indian LS data is considered invalid and is dropped, as the response scale changed between waves (Easterlin and Sawangfa 2010). This leaves 242 country-wave pairs.

The unit of analysis throughout the study is the country-wave pair. For each country-wave pair, we calculate the mean and Standard Deviation of LS (SD_{LS}). The latter is our measure of SWB inequality.[3] Figure 8.1b presents a histogram of the standard deviation of LS by country-wave pair. For each country-wave pair, we also calculate the percentage of respondents who are female, married, not parents, unemployed, and did not complete high school as well as the mean age.

The GDPpc data come from the World Bank's WDI. All GDPpc figures are in 2000 U.S. dollars. Of the 242 country-wave pairs in the analysis, there are GDPpc data for 237. GDPpc data are missing for Northern Ireland 1981, 1990, and 1999, and Taiwan 1994 and 2006; the World Bank does not recognize the countries for political reasons. We drop these country-wave pairs.

Finally, given that we are studying SWB inequality and that SWB inequality might be related to income inequality, we attempt to collect data regarding income inequality for the country-wave pairs in our data set. There is income

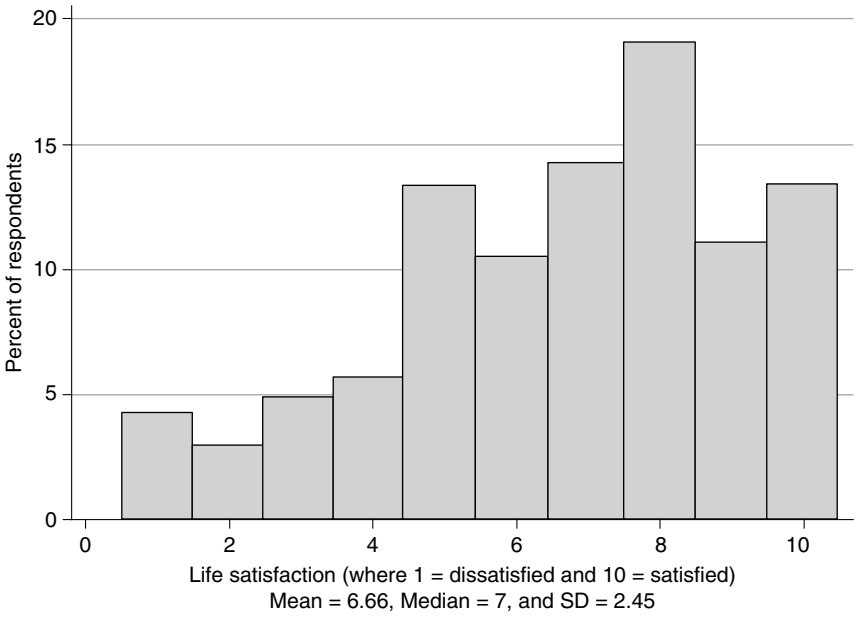


Figure 8.1a Distribution of LS responses in WVS (341,198 observations)

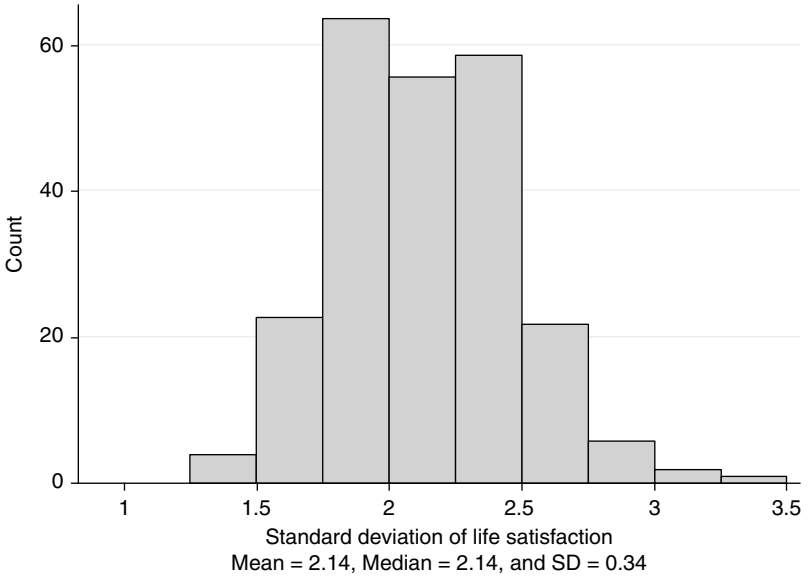


Figure 8.1b Distribution of standard deviation of LS by country-wave pair (237 observations)

inequality data (Gini coefficient) in the WDI for only 127 of the country-wave pairs, and we could not find a more complete source of income inequality data than the WDI. Thus, we use the Gini coefficient to measure income inequality only as a robustness check. As our primary measure of income inequality, we use, for each country-wave pair, the standard deviation of WVS respondents' self-reported income decile into which their household falls. We have this measure of income inequality for 227 country-wave pairs; it is missing for the following pairs: Argentina 1984 & 2006; Finland 1981; Hungary 1982 & 1998; Jordan 2007; Philippines 1996; Portugal 1999; Slovenia 1995; and Sweden 1990.

Table 8.1 presents the countries that are included in the data set, sorted by the average SD_{LS} across all the waves; it also shows for each country the number of WVS waves that were administered, the first and last year the WVS was administered, and the mean of LS and GDPpc across all the waves. Interestingly, the vast majority of the countries with the lowest (highest) SD_{LS} have high (low) GDPpc; the mean GDPpc of the 20 countries with the smallest (greatest) SD_{LS} is above $20,000 (below $2,000). Pakistan is a clear exception to this pattern; it has the lowest SD_{LS}, 1.46, and has an average GDPpc of $526. Further, the SD_{LS} decreases (and GDPpc increases), as one progressively restricts the sample to OECD countries to country-wave pairs with GDPpc greater than $10,000 and $20,000; Table 8.2 presents the mean SD_{LS}, LS, and other characteristics for country-wave pairs).

Table 8.1 Countries in WVS sorted by the standard deviation of life satisfaction (n=97)

Country	**Administered**			**Across all wave**		
	# waves	**First year**	**Last year**	**Mean LS**	**SD of LS**	**GDPpc (2000usd)**
Pakistan^	2	1997	2001	4.85	1.46	526
Netherlands^^^^	4	1981	2006	7.77	1.48	20,986
Iceland^^^^	3	1984	1999	8.04	1.60	26,674
Andorra^^^^	1	2005	2005	7.14	1.62	20,783
Finland^^^^	5	1981	2005	7.81	1.65	21,002
Switzerland^^^^	3	1989	2007	8.10	1.73	34,130
Sweden^^^^	5	1982	2006	7.82	1.75	25,133
Norway^^^^	4	1982	2008	7.80	1.75	31,561
Canada^^^^	4	1982	2006	7.82	1.77	21,370
Malaysia	1	2006	2006	6.84	1.79	4,792
Singapore^^^^	1	2002	2002	7.24	1.80	22,571
Thailand^^^^	1	2007	2007	7.21	1.81	2,592
Australia^^^^	3	1981	2005	7.59	1.81	18,623
Northern Ireland^^, ^^^^	3	1981	1999	7.85	1.82	–

(*continued*)

Table 8.1 Continued

Country	Administered			Across all wave		
	# waves	First year	Last year	Mean LS	SD of LS	GDPpc (2000usd)
Malta^^^^	3	1983	1999	8.15	1.84	7,079
Denmark^^^^	3	1981	1999	8.21	1.85	24,239
Spain^^^^	5	1981	2007	6.94	1.86	12,468
Japan^^^^	5	1981	2005	6.64	1.86	33,814
Luxembourg	1	1999	1999	7.81	1.87	43,421
United Kingdom^^^^	5	1981	2006	7.52	1.87	22,304
United States^^^^	5	1982	2006	7.60	1.88	30,375
Austria^^^^	2	1990	1999	7.95	1.88	21,188
Ireland^^^^	3	1981	1999	7.96	1.88	15,894
Hong Kong^^^^	1	2005	2005	6.41	1.93	30,395
Germany^^^^	5	1981	2006	7.10	1.94	20,719
Belgium^^^^	3	1981	1999	7.47	1.94	18,745
New Zealand^^^^	2	1998	2004	7.80	1.95	13,513
Colombia	2	1997	2005	8.31	1.97	2,579
Vietnam	2	2001	2006	6.81	1.98	500
France^^^^	4	1981	2006	6.84	1.99	19,739
Ethiopia^^^^	1	2007	2007	4.99	2.01	175
Taiwan^^, ^^^^	2	1994	2006	6.61	2.02	–
Cyprus^^^^	1	2006	2006	7.35	2.03	14,719
Albania	2	1998	2002	4.97	2.03	$1,167
Portugal^^^^	2	1990	1999	7.05	2.05	9,609
Czech Republic^^^,^^^^	3	1990	1999	6.72	2.06	5,301
Uruguay	2	1996	2006	7.30	2.07	7,127
Guatemala	1	2005	2005	7.95	2.09	1,762
Puerto Rico^^^^	2	1995	2001	8.30	2.09	15,178
Italy^^^^	4	1981	2005	7.00	2.10	16,971
Mexico^^^^	5	1981	2005	7.86	2.10	5,515
Indonesia^^^^	2	2001	2006	6.93	2.11	905
Slovenia^^^	4	1992	2005	6.81	2.11	9,184
Rwanda^^^^	1	2007	2007	4.97	2.11	290
Argentina	5	1984	2006	7.19	2.13	7,348
Chile	4	1990	2005	7.21	2.14	4,530
Morocco	2	2001	2007	5.66	2.17	1,499
Israel	1	2001	2001	7.03	2.17	19,366
Burkina Faso^^^^	1	2007	2007	5.57	2.18	260
Greece	1	1999	1999	6.67	2.19	11,043
Estonia^^^	3	1990	1999	5.64	2.20	3,535
Korea (South)^, ^^^^	5	1982	2005	6.16	2.20	9,247
Bangladesh	2	1996	2002	6.09	2.21	324
Belarus^^^	3	1990	2000	4.89	2.22	1,211
Croatia	2	1996	1999	6.43	2.22	4,421
Trinidad And Tobago^^^^	1	2006	2006	7.26	2.23	10,217
Saudi Arabia^^^^	1	2003	2003	7.28	2.27	9,266
Azerbaijan	1	1997	1997	5.39	2.29	513

(*continued*)

Table 8.1 Continued

Country	Administered			Across all wave		
	# waves	First year	Last year	Mean LS	SD of LS	GDPpc (2000usd)
Slovak Republic^^^, ^^^^	3	1990	1999	6.24	2.29	5,236
Bosnia And Herzegovina^^^^	2	1998	2001	5.61	2.30	1,407
Serbia^^^^	3	1996	2006	5.77	2.31	1,368
Poland^^, ^^^	4	1989	2005	6.55	2.32	4,112
Moldova	3	1996	2006	4.58	2.33	427
China^^^, ^^^^	4	1990	2007	6.85	2.35	970
Latvia^^^	3	1990	1999	5.29	2.35	3,148
Peru	3	1996	2008	6.61	2.35	2,309
Armenia	1	1997	1997	4.32	2.37	520
Hungary^^^, ^^^^	4	1982	1999	6.15	2.39	4,103
Ukraine	3	1996	2006	4.77	2.39	747
Brazil	3	1991	2006	7.39	2.40	3,712
Iran^^^^	2	2000	2007	6.40	2.41	1,861
Bulgaria^^^	4	1990	2006	5.10	2.41	1,697
Iraq^^^^	2	2004	2006	4.84	2.41	711
El Salvador	1	1999	1999	7.50	2.43	2,174
Georgia	2	1996	2008	4.82	2.43	891
Philippines	2	1996	2001	6.75	2.44	951
Macedonia	2	1998	2001	5.41	2.45	1,673
Dominican Republic	1	1996	1996	7.13	2.47	2,227
Uganda	1	2001	2001	5.65	2.47	258
Russian Federation^^^, ^^^^	4	1990	2006	5.16	2.48	2,120
Romania^^^	4	1993	2005	5.43	2.49	1,767
Zambia^^^^	1	2007	2007	6.06	2.50	374
Turkey^^^^	4	1990	2007	6.41	2.50	3,992
Nigeria	3	1990	2000	6.68	2.52	362
Lithuania^^^	3	1990	1999	5.40	2.54	3,458
South Africa^^^^	5	1982	2007	6.62	2.56	3,279
Kyrgyz Republic	1	2003	2003	6.48	2.57	306
Mali	1	2007	2007	6.09	2.59	287
Ghana	1	2007	2007	6.12	2.63	313
Jordan	2	2001	2007	6.40	2.65	2,091
Venezuela	2	1996	2000	7.12	2.75	4,912
Zimbabwe^^^^	1	2001	2001	3.95	2.79	576
Algeria	1	2002	2002	5.67	2.86	1,874
Egypt^^^^	2	2000	2008	5.57	3.02	1,604
Tanzania	1	2001	2001	3.87	3.22	283

^Missing LS data from WVS: Korea 1996; and Pakistan 1997.
^^Missing GDPpc from WDI: Northern Ireland 1981, 1990, & 1999; and Taiwan 1994 & 2006. Poland 1989 GDPpc data from 1990 (1989 data missing).
^^^Transition country.
^^^^Missing Gini coefficient from WDI for at least one wave.

Table 8.2 Mean characteristics by country-wave pairs

	All (1)		OECD (2)		GDPpc > $10,000 (3)		GDPpc > $20,000 (4)	
Mean LS	6.69	(0.07)	7.23	(0.06)	7.48	(0.05)	7.55	(0.07)
SD_{LS}	2.14	(0.02)	1.96	(0.02)	1.85	(0.02)	1.80	(0.02)
GDPpc (in 2000usd)	$10,283	(691)	$17,154	(945)	$22,201	(813)	$27,509	(848)
Income inequality+	2.22	(0.03)	2.37	(0.04)	2.41	(0.04)	2.50	(0.06)
Age	42.07	(0.39)	44.00	(0.47)	44.56	(0.55)	45.73	(0.88)
Female	0.52	(0.00)	0.52	(0.00)	0.53	(0.00)	0.52	(0.00)
Married	0.58	(0.01)	0.58	(0.01)	0.57	(0.01)	0.56	(0.01)
No children	0.25	(0.01)	0.23	(0.01)	0.25	(0.01)	0.27	(0.02)
Unemployed	0.08	(0.00)	0.05	(0.00)	0.05	(0.00)	0.05	(0.00)
Did not complete high school	0.45	(0.01)	0.45	(0.02)	0.41	(0.03)	0.43	(0.03)
Number of countries	93		34		32		21	
Number of country-wave pairs	237		120		90		50	

Standard error in parenthesis.
+Standard deviation of reported income deciles

8.4 Results

To examine the relationship between SWB inequality and income we first treat the data as repeated cross-sections. This analysis provides strong evidence that the two are negatively correlated, indicating that countries with higher income have lower SWB inequality. Next, we treat the data as time series. This analysis provides some evidence that countries with the greatest economic growth rates experience the greatest decrease in SWB inequality. However, the time series analysis is far from conclusive.

A Cross-section analysis

Figure 8.2a plots the SD_{LS} and the Natural Log of GDPpc (LGDPpc) for each country-wave pair. There appears to be a negative relationship. That is, SD_{LS} is smaller in country-wave pairs with greater LGDPpc (Figure 8.2b illustrates that the relationship is similar but less linear when one compares SD_{LS} and GDPpc). Given that log income is generally used when studying the relationship between mean SWB and income, we use LGDPpc in the subsequent analysis, unless noted otherwise.

To estimate the relationship between SWB inequality and income, an equation of the following form is estimated:

$$SD^{LS}_{c-w} = \alpha\, LGDPpc_{c-w} + \beta \bullet X_{c-w} + \varepsilon_{c-w} \tag{1}$$

where SD^{LS}_{c-w} is the SD_{LS} for each country-wave pair, $c-w$; $LGDPpc_{c-w}$ is the natural log of GDPpc in 2000 US dollars for each $c-w$ pair; and X_{c-w} is a matrix of

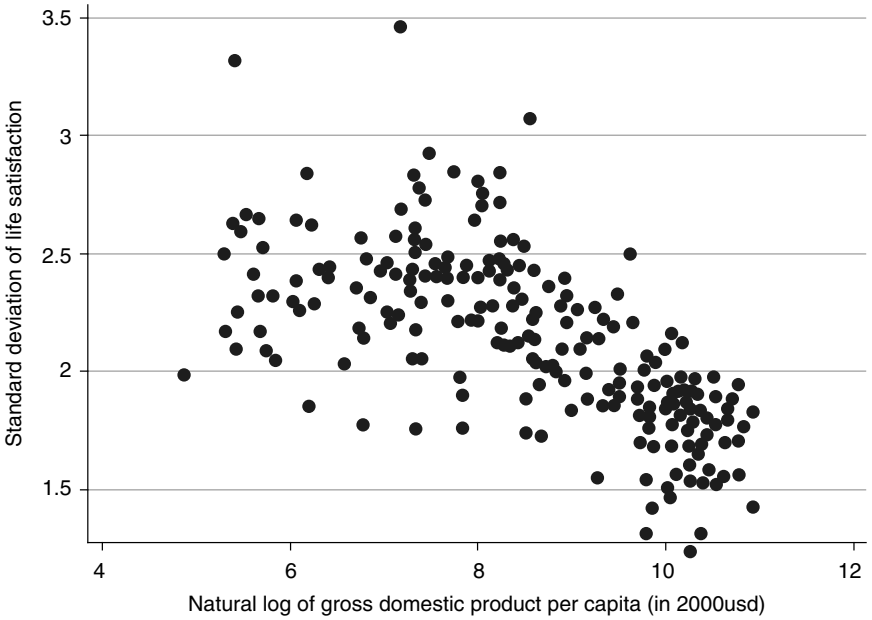


Figure 8.2a Scatterplot of standard deviation of life satisfaction and natural log of per capita GDP (in 2000USD)

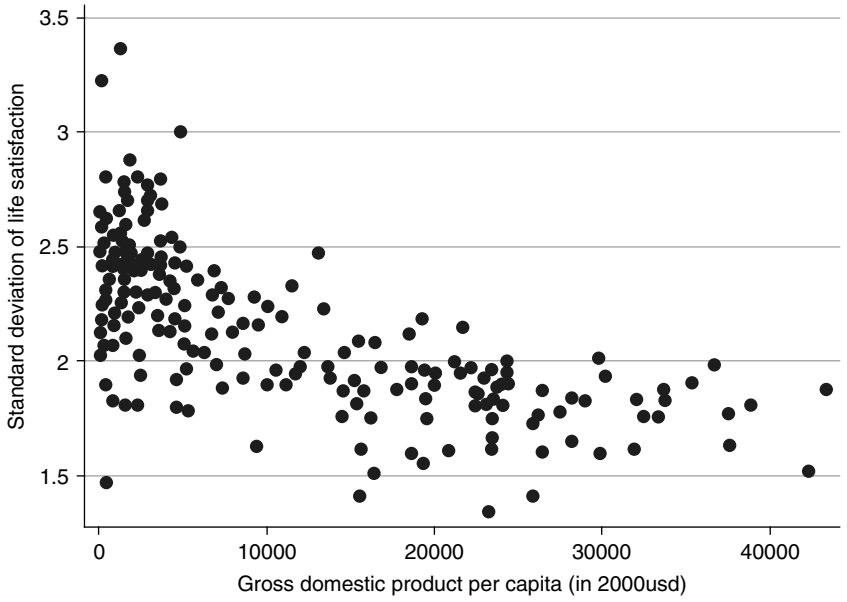


Figure 8.2b Scatterplot of standard deviation of life satisfaction and gross domestic product per capita (in 2000USD)

characteristics for each $c-w$ pair including mean LS, income inequality, mean age, and percent of respondents who are female, married, childless, unemployed, and not high school graduates.[4] Equation (1) is estimated using OLS and country fixed effects.

Estimating equation (1) without covariates, the coefficient on LGDPpc is negative and highly statistically significant, confirming the negative relationship between SD_{LS} and LGDPpc that is apparent in Figure 8.2a (Column 1 of Table 8.3). Column 2 shows that this finding is robust to adding income inequality (using the SD of reported income), mean LS, and the other regressors discussed above (results from the progressive addition of these regressors are shown in Table 8.4). The magnitude of the coefficient indicates that doubling GDPpc is associated with a 0.19 reduction in SD_{LS}, or a 9 percent (=0.19/2.14) reduction from mean SD_{LS}. This is equivalent to moving from 46th (Chile, SD_{LS}=2.14) to 27th (New Zealand, SD_{LS}=1.95) in the SWB inequality ranking. The coefficient on income inequality is positive but statistically insignificant, which indicates that the negative relationship between SD_{LS} and GDPpc is not simply an artifact of a negative relationship between GDPpc and income inequality. Finally, the coefficient on mean LS is negative and statistically significant, indicating that as mean LS increases, SD_{LS} decreases. These results are not driven by the transition economies: Columns 3 and 4 of Table 8.3 show that the result is robust to, and indeed strengthened by, the exclusion of transition countries.

Table 8.3 Ordinary least square estimates of equation (1) where the dependent variable is the standard deviation of life satisfaction

	All (1)	**All (2)**	**Non-transition countries (3)**	**Non-transition countries (4)**
LGDPpc	–0.187***	–0.192**	–0.255***	–0.351***
	(0.058)	(0.082)	(0.072)	(0.116)
Income inequality[+]		0.024		0.019
		(0.033)		(0.042)
Mean LS		–0.085**		–0.086*
		(0.040)		(0.052)
Includes:				
Country fixed effects	Yes	Yes	Yes	Yes
Other covariates[++]	No	Yes	No	Yes
Observations	237	237	191	191
Number of countries	93	93	80	80

Standard errors in parenthesis.
*, **, ***signifies p<0.10, 0.05, 0.01, respectively.
[+]For each country-wave pair, the standard deviation of income.
[++]For each country-wave pair, mean age; percent of respondents did not complete high school; and percent of respondents who are female, married, not parents, and unemployed.

Table 8.4 Ordinary least square estimates of equation (1) where the dependent variable is the standard deviation of life satisfaction

	All (1)	All (2)	All (3)	All (4)	All (5)	Non-transition countries (6)	Non-transition countries (7)
Log GDP per capita	–0.157***	–0.187***	–0.181***	–0.130**	–0.192**	–0.255***	–0.351***
	(0.016)	(0.058)	(0.058)	(0.061)	(0.082)	(0.072)	(0.116)
Income inequality[+]			0.045	0.044	0.024		0.019
			(0.033)	(0.032)	(0.033)		(0.042)
Mean LS				–0.084**	–0.085**		–0.086*
				(0.037)	(0.040)		(0.052)
Includes:							
Country fixed effects	No	Yes	Yes	Yes	Yes	Yes	Yes
Other covariates[++]	No	No	No	No	Yes	No	Yes
Observations	237	237	237	237	237	191	191
Number of countries	93	93	93	93	93	80	80

Standard errors in parenthesis.
*, **, *** signifies $p<0.10$, 0.05, 0.01, respectively.
[+]For each country-wave pair, the standard deviation of income.
[++]For each country-wave pair, mean age; percent of respondents did not complete high school; and percent of respondents who are female, married, not parents, and unemployed.

The above analysis is repeated, using the 127 country-wave pairs with Gini information. Table 8.5 compares the coefficients on LGDPpc and income inequality using as measures of income inequality the SD of reported income (as in Table 8.3) and the country-wave Gini coefficient. While the restricted sample entails a loss of statistical power when other covariates are added (Columns 4 and 5 of Table 8.5), the coefficient on LGDPpc is stable across specifications (ranging from –0.225 to –0.314) and statistically significant in the absence of covariates (Columns 2 and 3). Thus the results appear robust to the choice of income-inequality metric.

To determine whether the decrease in SD_{LS} associated with greater GDPpc results from fewer reports of "low LS" or "high LS," we estimate equation (1) with corresponding binary variables in place of SD^{LS}_{c-w}. Specifically, *Low LS*, equals one if LS equals 1, 2, 3, or 4, and zero otherwise, and *High LS* equals one if LS is 9 or 10, and zero otherwise. Higher GDPpc is associated with a statistically significant reduction in low LS. With country fixed effects and no other covariates, doubling GDPpc is associated with a 9.2 percentage point reduction in a respondent's likelihood of reporting low LS (Column 1 of Table 8.6). The corresponding specification shows no statistically significant relationship between GDPpc and high LS (Column 4). Next, in columns 2 and 5, we include the controls described above; importantly, these include mean LS, which is known to be a positive correlate of income in cross-sectional analysis. A noteworthy result emerges: now, higher GDPpc is associated with a statistically significant decrease in both low and high LS (Columns 2 and 5).

Table 8.5 Ordinary least square estimates of equation (1) where the dependent variable is the standard deviation of life satisfaction, using WDI data and comparing the standard deviation of reported income to the Gini coefficient

	(1)	**(2)**	**(3)**	**(4)**	**(5)**
LGDPpc	–0.305***	–0.290**	–0.314**	–0.251	–0.225
	(0.118)	(0.122)	(0.121)	(0.182)	(0.168)
SD of reported income		0.026		–0.029	
		(0.051)		(0.063)	
Gini coefficient			0.003		0.001
			(0.007)		(0.009)
Includes:					
Country fixed effects	Yes	Yes	Yes	Yes	Yes
Other covariates[+]	No	No	No	Yes	Yes
Observations	127	127	127	127	127
Number of countries	69	69	69	69	69

Standard errors in parenthesis.
*,**,***signifies p<0.10, 0.05, 0.01, respectively.
[+]For each country-wave pair, mean LS, mean age, and percent of respondents who are female, married, not parents, unemployed, and not high school graduates.

Table 8.6 Ordinary least square estimates of equation (1) where the dependent variable is low and high levels of life satisfaction

	Low LS (1)	**Low LS (2)**	**Low LS Non-transition countries (3)**	**High LS (4)**	**High LS (5)**	**High LS Non-transition countries (6)**
LGDPpc	–0.092***	–0.026**	–0.032**	0.015	–0.060***	–0.087***
	(0.018)	(0.011)	(0.015)	(0.019)	(0.019)	(0.028)
Income inequality[+]		0.009*	0.007		0.008	0.010
		(0.005)	(0.006)		(0.008)	(0.010)
Mean LS		–0.124***	–0.113***		0.106***	0.120***
		(0.005)	(0.007)		(0.009)	(0.013)
Includes:						
Country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Other covariates[++]	No	Yes	Yes	No	No	Yes
Observations	237	237	191	237	237	191
Number of countries	93	93	80	93	93	80

Standard errors in parenthesis.
*, **, ***signifies $p<0.10$, 0.05, 0.01, respectively.
[+]For each country-wave pair, the standard deviation of income.
[++]For each country-wave pair, mean age; percent of respondents did not complete high school; and percent of respondents who are female, married, not parents, and unemployed.

Interestingly, as GDPpc increases, high LS decreases, indicating that after controlling for mean LS, high LS actually decreases with GDPpc. Excluding transition countries increases the magnitude of the LGDPpc coefficient (Columns 3 and 6), indicating that the pooled results are not driven by these countries.

Finally, as shown in Table 8.7, restricting the sample to more developed countries, either the OECD countries or those countries with GDPpc greater than $10,000 or $20,000, increases the magnitude of the coefficient on LGDPpc. This indicates that the negative relationship between SD_{LS} and GDPpc is more negative for higher-income countries.

B Time series analysis

The negative relationship between SWB inequality and income that is apparent in cross-sectional analysis may or may not persist in time series analysis. To investigate we examine the evolution over time of SD_{LS} and GDPpc in individual countries. First we focus on countries with the longest time series, that is, the ten countries for which we have fives waves of LS data, providing a 22+ year time series for each: Argentina, Finland, Germany, Japan, Mexico, South Africa, Spain, Sweden, the United Kingdom, and the United States. Of these countries, all experience weakly increasing GDPpc over the time period. Figure 8.3 presents the time series of GDPpc and SD_{LS} for the United States.

Table 8.7 Ordinary least square estimates of equation (1) where the dependent variable is the standard deviation of life satisfaction and sample is limited to OECD countries and also countries with high GDP per capita

	OECD (1)	**OECD (2)**	**Per capita income > $10,000 (3)**	**Per capita income > $10,000 (4)**	**Per capita income > $20,000 (5)**	**Per capita income > $20,000 (6)**
Log GDP per capita	–0.200***	–0.318**	–0.214**	–0.448**	–0.388***	–0.530***
	(0.066)	(0.121)	(0.086)	(0.157)	(0.101)	(0.171)
Income inequality[+]		0.008		0.012		0.044
		(0.037)		(0.049)		(0.051)
Mean LS		–0.156***		–0.299***		–0.222*
		(0.048)		(0.096)		(0.109)
Includes:						
Country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Other covariates[++]	No	Yes	No	Yes	No	Yes
Observations	120	120	90	90	50	50
Number of countries	34	34	32	32	21	21

Standard errors in parenthesis.
*, **, *** signifies $p<0.10$, 0.05, 0.01, respectively.
[+]For each country-wave pair, the standard deviation of income.
[++]For each country-wave pair, mean age; percent of respondents did not complete high school; and percent of respondents who are female, married, not parents, and unemployed.

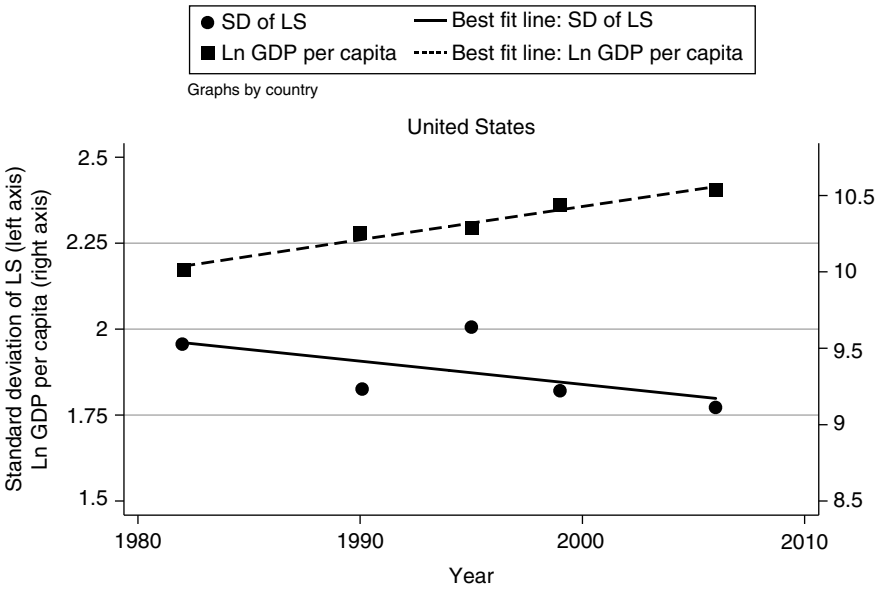


Figure 8.3 Time series of US standard deviation of life satisfaction and the natural log of per capita GDP, 1982–2006

We calculate the average annual percentage change in the SD_{LS} and GDPpc between the first and last observation for each country. For example,

$$Avg\%\Delta SD_c^{LS} = \left(\frac{SD_{c,w=5}^{LS} - SD_{c,w=1}^{LS}}{SD_{c,w=1}^{LS}} \right) \Big/ years_c \qquad (2)$$

where $year_c$ is the number of years country, c, is in the time series. Figure 8.4 presents the scatterplot of these calculations for the ten countries that administered the WVS in all five waves. There appears to be a negative relationship, indicating that countries that experience the greatest average per-capita growth rate experience the greatest reduction in SD_{LS}.

Regressing $Avg\%\Delta SD_c^{LS}$ on $Avg\%\Delta GDPpc_c$, one finds a negative but statistically insignificant relationship (Column 1 of Table 8.8). However, the number of observations is small and there is one clear outlier, Finland (see Figure 8.4). Dropping Finland, one finds that the coefficient remains negative, grows in magnitude, and is statistically significant; this holds in the simple regression and with controls for the average annual percentage change in SD of reported income and mean LS (Columns 2 and 3). In summary, there is evidence that the countries that experience the greatest per capita economic growth also

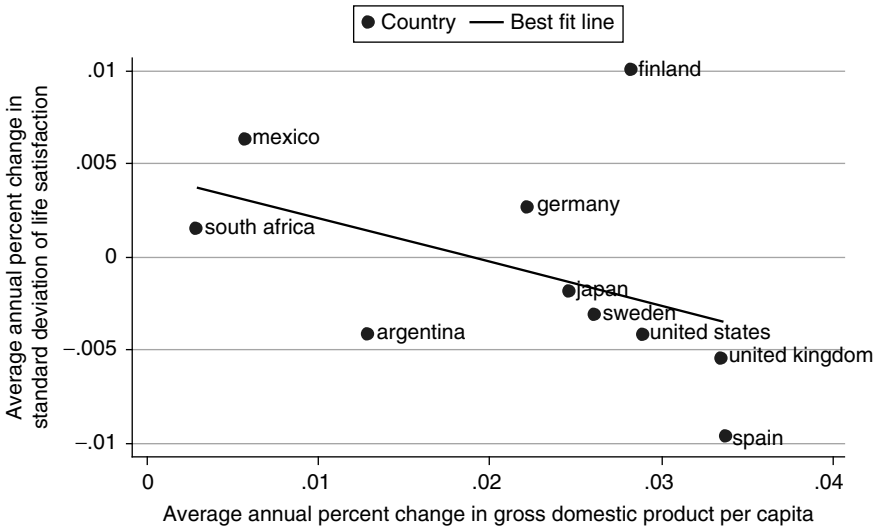


Figure 8.4 Scatterplot of average annual percent change in standard deviation of life satisfaction and average annual percent change in per capita GDP for countries in five waves of WVS

Table 8.8 Ordinary least squares estimates of regressing the average annual change in the standard deviation of life satisfaction on average annual per capita GDP growth

	In 5 waves (1)	**In 5 waves, except Finland (2)**	**In 5 waves, except Finland (3)**	**In 4+ waves, except China and Korea (4)**	**In 4+ waves, except China, and Korea (5)**
Avg%ΔGDPpc	–0.235	–0.319**	–0.315**	–0.164***	–0.181***
	(0.142)	(0.099)	(0.118)	(0.046)	(0.049)
Avg%ΔIncome			0.306		0.054
Inequality			(0.247)		(0.034)
Avg%ΔMean of LS			–0.875*		–0.414**
			(0.391)		(0.169)
Number of countries	10	9^	9^	23	21^

Standard errors in parenthesis.
*,**,*** signifies p< 0.10, 0.05, 0.01, respectively.
+For each country-wave pair, percent change from first to last wave.
^ Argentina (in 5 waves) and Hungary (in 4 waves) are dropped due to missing income inequality data.

experience the greatest decrease in SD_{LS}. The magnitude of the coefficient indicates that if GDPpc doubles, then SD_{LS} will decrease by 20–30 percent. This is equivalent to moving from 46th (Chile, SD_{LS}=2.14) to 6th (Switzerland, SD_{LS}=1.73) in the SWB inequality ranking.

Broadening the analysis to include countries with at least four waves of LS data provides a 12+ year time series for 25 countries (additional countries

include Bulgaria, Canada, Chile, China, France, Hungary, Italy, Korea (South), the Netherlands, Norway, Poland, Romania, Russia, Slovenia, and Turkey). Now there is a positive relationship between $Avg\%\Delta SD_c^{LS}$ and $Avg\%\Delta GDPpc_c$, as illustrated with the solid line in Figure 8.5. However, there are two outliers, China and Korea, whose growth rates are each more than twice as large as the next fastest-growing economies. Dropping the greatest outlier, China, from the figure materially changes the best-fit relationship (long-dashed line, Figure 8.5) to a negative one. Further, if one drops Korea, the country with the next fastest growth rate, then the negative relationship becomes greater (short-dashed line, Figure 8.5). In the next section, we briefly discuss why unusually high growth rates may be associated with increased SD_{LS}.

Regressing $Avg\%\Delta SD_c^{LS}$ on $Avg\%\Delta GDPpc_c$ for the countries in at least four waves of the WVS, one finds negative and statistically significant relationship with China and Korea excluded (Column 4 of Table 8.8). This result holds with the inclusion of controls for the average annual percent changes in both SD of reported income and mean LS. That is, excluding the two countries with the greatest economic growth rates, it appears that countries experiencing greater economic growth also experience greater decreases in SD_{LS}. An alternative explanation, for which we have no statistically significant support,

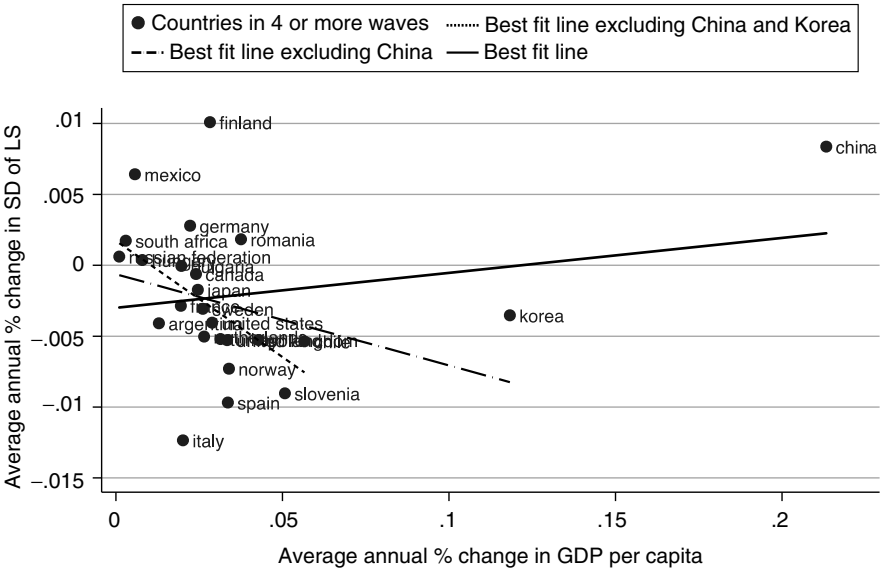


Figure 8.5 Scatterplot of average annual percent change in standard deviation of life satisfaction and average annual percent change in per capita GDP for countries in at least four waves of WVS

is that, in time series, there is a U-shaped relationship between economic growth and changes in SD_{LS}, wherein China and Korea lie on the upward-sloping part of the U.

8.6 Discussion

We present evidence that there is a negative relationship between SWB inequality and income in cross-sectional analysis; this result is stable regardless of the covariates included in the analysis. The results indicate that the doubling of income is associated with a 9 percent reduction in SD_{LS}, our measure of SWB inequality, from the mean. There is also time series evidence that for most countries, greater economic growth rates will also be associated with greater declines in SWB inequality over time. The results indicate that doubling income is associated with a 20–30 percent decrease in SWB inequality. Interestingly, this pattern is contradicted for the two countries in the data set with the greatest economic growth rates: China and Korea. Perhaps, exceptional economic growth rates do not lead to decreasing SWB inequality over time, as such growth rates might cause large changes that affect citizens' SWB in disparate ways. Such a relationship is corroborated for China in Easterlin et al. (2012), which documents that those people in the bottom third of the income distribution were the most hard-hit by the reduced job security, and associated benefits thereof, entailed by SOE restructuring.

The decrease in SWB inequality associated with economic growth seems to be associated with a decrease in low LS. This contrasts sharply with the recent positive correlation of economic growth and income inequality. For example, the United States and the United Kingdom have each experienced well-documented increases in income inequality during recent periods of economic growth. In contrast, economic growth appears to be negatively associated with high LS. The investigation of why greater income is associated with a compression of the LS distribution – for example, hedonic adaptation (Di Tella, Haisken-De New, and MacCulloch 2010) and negative side effects of attaining increased income on the high-LS end of the distribution, and improved social safety nets at the low-LS end of the distribution – is left for future research. To this end, Easterlin (1995) illustrates the relationship between social safety nets and mean LS.

Because our results are for the most part consistent across cross-sectional and time series analyses, they do not present the challenge that the Easterlin paradox does. Our research also suggests that, despite the controversy the Easterlin paradox presents, there may be an additional benefit – insofar as SWB equality is desirable – associated with increased per capita income within a country: namely, decreasing SWB inequality.

Notes

1. Clark, Flèche, and Senik (2012) was developed simultaneously and independently from this paper, on the same topic.
2. This article was published in a special issue of the Journal of Happiness Studies (volume 6, number 4), which includes four papers that focus on happiness inequality (Veenhoven 2005a; Kalmijn and Veenhoven 2005; Veenhoven and Kalmijn 2005; and Ott 2005). Veenhoven (2005b) introduces the series. The other three papers are more pertinent to methodology, one of which we cite in the data section.
3. Kalmijn and Veenhoven (2005) compare the effectiveness of eight SWB inequality measures, including a Gini coefficient, standard deviation, absolute mean difference, and inter-quartile range. After examining each statistic and empirically testing their sensitivity to various distributions, they determine that four statistics are adequate measures of SWB inequality, one of which is the standard deviation. Since the standard deviation is widely used and understood, we use it to measure SWB inequality.
4. To reduce omitted variable bias, it is standard to include these demographic controls when regressing mean-SWB on income, as they are well-documented correlates of SWB. Their inclusion here is to ensure that they are not driving any observed correlation between SWB-inequality and income. Excepting "percent married" in some specifications, they are not statistically significant determinants of SWB-inequality. The small existing literature on SWB-inequality often includes such demographic controls, but, to the authors' knowledge, nowhere has there been a systematic analysis of the relationship between them and SWB-inequality, and such an analysis is outside of the scope of this paper. Importantly, the paper's main results are robust to the inclusion of these controls.

References

Atkinson, A.B. (1970) "On the Measurement of Inequality," *Journal of Economic Theory*, vol. 2, pp. 244–63.

Clark, A.E., S. Flèche and C. Senik (2012) "The Great Happiness Moderation." Paper presented at the CEPREMAP-PSE-DIMeco Conference Happiness and Economic Growth: Lessons from Developing Countries, Paris, March 9.

Diener E. and M.E.P. Seligman (2004) "Beyond Money: Toward an Economy of Well-being," *Psychol Sci Public Interest*, vol. 5, pp. 1–31.

Di Tella, R.M., J. Haisken-De New and R. MacCulloch (2010) "Happiness Adaptation to Income and to Status in an Individual Panel," *J Econ Behav Organ*, vol. 76, pp. 834–52.

Easterlin, R.A. (1974) in P.A. David and M.W. Reder (eds), *Nations and Households in Economic Growth: Essays in Honor of Moses Abramowitz*. New York: Academic Press, pp. 89–125.

Easterlin, R.A. (1995) "Will Raising the Incomes of All Increase the Happiness of All?," *J Econ Behav Organ*, vol. 27, pp. 35–47.

Easterlin, R.A. (2012) "Life Satisfaction of Rich and Poor Under Socialism and Capitalism," *International Journal of Happiness and Development*, vol. 1, pp. 112–26.

Easterlin, R.A. (2013) "Happiness, Growth, and Public Policy, 2 *Econ Inq*, vol. 51, pp. 1–15.

Easterlin, R.A., L.A. McVey, M. Switek, O. Sawangfa and J.S. Zweig (2010) "The Happiness–Income Paradox Revisited," *Proc Natl Acad Sci U S A*, vol. 107, pp. 22463–8.

Easterlin, R.A., R. Morgan, M. Switek and F. Wang (2012) "China's Life Satisfaction: 1990–2010," *Proc Natl Acad Sci U S* A, vol. 109, pp. 9775–80.

Easterlin, R.A. and O. Sawangfa (2010) "Happiness and Economic Growth: Does the Cross-section Predict Time Trends? Evidence from Developing Countries," in E. Diener, J. Helliwell and D. Kahneman (eds), *International Differences in Well-Being*. Princeton, NJ: Princeton University Press, pp. 162–212.

Gastwirth, J.L. (1972) "The Estimation of the Lorenz Curve and Gini Index," *Rev Econ Stat*, vol. 54, pp. 306–16.

Gini, C (1921) "Measurement of Inequality of Incomes," *Econ J*, vol. 31, pp. 124–6.

Gottschalk, P. and T. Smeeding (1997) "Cross-national Comparisons of Earnings and Income Inequality," *J Econ Lit*, vol. 35, pp. 633–88.

Kahneman, D. and A.B. Krueger (2006) "Developments in the Measurement of Subjective Well-being," *J Econ Perspect*, vol. 20, pp. 3–24.

Kalmijn, W. and R. Veenhoven (2005) "Measuring Inequality of Happiness in Nations: In Search for Proper Statistics," *J Happiness Stud*, vol. 6, pp. 357–96.

Kuznets, S. (1955) "Economic Growth and Income Inequality," *Am Econ Rev*, vol. 45, pp. 1–28.

Lorenz, M.O. (1905) "Methods of Measuring the Concentration of Wealth," *Publications of the American Statistical* Association, vol. 9, pp. 209–19.

Ott, J. (2005) "Level and Inequality of Happiness in Nations: Does Greater Happiness of a Greater Number Imply Greater Inequality in Happiness?," *J Happiness Stud*, vol. 6, pp. 397–420.

Stevenson, B. and J. Wolfers (2008a) "Economic Growth and Subjective Well-being: Reassessing the Easterlin Paradox," *Brookings Pap Econ* Act, vol. 39, pp. 1–102.

Stevenson, B. and J. Wolfers (2008b) "Happiness Inequality in the United States," *J Legal Stud*, vol. 37, pp. 533–79.

Stevenson, B. and J. Wolfers (2013) "Subjective Well-being and Income: Is There Any Evidence of Satiation?," *Am Econ Rev: Papers and Proceedings*, vol. 103, pp. 598–604.

Stiglitz, J.E., A. Sen, and J.-P. Fitoussi (2010) *Mis-Measuring Our Lives*. New York: The New Press.

Veenhoven, R (2005a) "Return of Inequality in Modern Society? Test by Dispersion of Life-satisfaction Across Time and Nations," *J Happiness Stud*, vol. 6, pp. 457–87.

Veenhoven, R (2005b) "Inequality of Happiness in Nations," *J Happiness Stud*, vol. 6, pp. 351–5.

Veenhoven, R. and W. Kalmijn (2005) "Inequality-adjusted Happiness in Nations: Egalitarianism and Utilitarianism Married in a New Index of Societal Performance," *J Happiness Stud*, vol. 6, pp. 421–55.

The World Bank *World Development Indicators*. Washignton, DC: The World Bank.

World Values Survey 1981–2008 Official Aggregate v.20090901, 2009. World Values Survey Association (www.worldvaluessurvey.org). Aggregate File Producer: ASEP/JDS, Madrid.

Why Focus on Subjective Wellbeing Inequality? Comments on "Inequality of Happiness: Evidence of the Compression of the Subjective-Wellbeing Distribution with Economic Growth" by John Ifcher and Homa Zarghamee

Murray Leibbrandt
University of Cape Town

Spawned by Easterlin, there is a large literature on the relationship between levels of income and levels of subjective wellbeing (SWB) and between economic growth and changes in SWB. This corpus reveals a positive relationship between levels of income and levels of SWB. But the positive relationship disappears when looking at the changes in income and changes in SWB. This is the Easterlin paradox and it has piqued the interest of the research community. Why? The positive relationship between levels of income and levels of SWB is unsurprising, but if it is very strong, it could be seen as supporting the idea that income and SWB are substitute measures of wellbeing or development. Such an idea would bewilder any microeconomist for whom income represents the budget constraint and SWB represents utility. To give one example of this framework in action: the work on the non-pecuniary costs of unemployment examines the relationship between SWB and unemployed, controlling for income. Such an idea would also bewilder development economists trained in the capabilities approach to development. Income is one enabler in achieving capabilities, albeit an important one in any market economy. Or, to give this statement its dynamic expression, GDP growth is only one component of development.

The fact that the cross-sectional and time series relationships reverse is, by definition, a paradox. But the fact that this affirms a complicated relationship, even at the macro-level, between wellbeing and income is unsurprising and even comforting to development economists. Indeed, perhaps the most useful aspect of this corpus has been in pushing on from these empirical findings to uncover the particularly important impactors of mean SWB and changes in SWB, controlling for mean income or for growth.

The paper under discussion seeks to make a particular contribution to this literature. It moves away from an interrogation of the drivers of mean SWB and

its change to a focus on the inequality of SWB and its change and how these relates to GDP or the growth of GDP. In line with the first part of the Easterlin literature, the authors find that a doubling of income is associated with a noticeable (9 percent) reduction from its mean in the standard deviation of SWB. Moreover, the negative relationship between the inequality of SWB and the level of per capita GDP (GDPpc) strengthens for higher-income countries. There might be evidence of non-linearities but, unlike the microeconomics literature on happiness, they are not indicative of diminishing SWB returns to GDPpc. Importantly, when turning to the time series estimates of growth, the Easterlin paradox does not manifest. Consistent with the cross-sectional evidence, positive growth is associated with declines in SWB inequality and greater growth rates are associated with greater annual average declines in SWB inequality.

The authors conclude in the following way:

> Because our results are for the most part consistent across these two analyses, they do not present the challenge that the Easterlin paradox does. Our research also suggests that, despite the controversy the Easterlin paradox presents, there may be an additional benefit from increasing per capita income within a country: namely, decreasing SWB inequality.

This is quite tame. Certainly, the authors cannot be accused of overstating the importance of their research. They are clear that they see their core contribution as an empirical one; namely, a careful, best-practice replication of the Easterlin corpus but examining inequality of SWB rather than mean SWB. They are successful in this endeavour. Is this valuable? The best-case scenario is one in which these stylized facts stimulate debates and a flourishing research agenda as happened with the Easterlin paradox.

But I feel that the authors could be more proactive in motivating the importance of this approach. For my taste, they are too narrow and understated in the way that they frame the paper and discuss the results. In terms of framing, we are missing a discussion right up front as to why we would want to put inequality of SWB on the left-hand side of an Easterlin-type model in either its cross-sectional or time series versions. There are alluring reasons. Surely an index of the diversity of SWB in a country is potentially a more inclusive or encompassing proxy for a society's wellbeing or experience of development than mean SWB? Earlier I referred to the importance of the Easterlin paradox in terms of its affirmation that development is more than growth. Mean SWB is unlikely to be adequate in capturing this texture of growth. The inequality of SWB would seem to be first-order dominant as an indicator capturing the fact that there are winners and losers in every growth process. Here one thinks about the Kuznets curve literature. The reason why development economists

have invested so much time looking at the relationship between income inequality and economic growth is that this provides a canvas on which to document and understand the economic transformations that unfold as part of the process of economic growth. One can argue that the relationship between the inequality of SWB and economic growth is an alluring canvas too.

Perhaps I overstate in order to compensate for the understatement by the authors. Nonetheless I am comfortable using my comments to strengthen the case that there is a point to all of this careful data work and that this is an interesting and potentially important line of inquiry. If one is prepared to dig around a little in the paper, a fair amount of interesting evidence is presented to add to this case.

The authors themselves give a cogent example to back up my comparison with the Kuznets curve debates. They remind us of the contrast between the negative correlation between growth and inequality of SWB and recent evidence from the globalized world of the positive correlation between growth and income inequality. These orthogonal findings make it clear that the inequality of SWB is picking up something substantively different from income inequality. The modelling work in the paper pushes this point further. Income inequality is included as a control variable in nearly all specifications. The most important point for the authors is the fact that this control does not change the sign or the significance of the coefficient on income or growth. For me it is at least as important that the coefficient on income inequality is statistically insignificant in all cases. At one level this seems implausible. One would think that socio-economic polarization would impact the inequality of SWB independently of mean income or of growth. But the data work is careful and the result cannot be set aside. We are prodded to think hard about what might be going on in the relationship between the inequality of SWB and the inequality of income controlling for income.

Aside from inequality, there are other interesting controls that are included in the cross-sectional estimates. These include mean SWB, age and the percentage of respondents who are female, married, childless, unemployed, and not high school graduates. Again, these controls are motivated as a standard set of controls from the SWB literature. However, the shift to a focus on the inequality of SWB is substantive and I would have preferred a stronger motivation in this context. At the technical level, it seems to me that mean SWB is crucial in ensuring that this analysis of the inequality of SWB is not unwittingly picking up level effects. More substantively, there are many interesting controls that resonate loosely with the empirical literature on the microeconomics of happiness and seem to belong in an equation explaining the inequality of SWB. But do they belong and, aside from the fact that they have no impact on the income coefficient, what can we learn from their estimated effects? Perhaps their impact is mopped up already by the inclusion of the level of SWB on the right-hand side?

Right up front the authors review the thin preceding literature on the inequality of SWB. It is clear that this literature goes down this path because it is seen to add an additional dimension to the standard Easterlin framework. Veenhoven (2005) uses the inequality of SWB to argue for the end of the great U-turn and the return of social inequality in modern society. Then, Easterlin (2012) himself compares the inequality of SWB in capitalist societies and socialist societies before and after they transition to capitalism. Pre-transition, the capitalist societies have bigger SWB inequality, post-transition the socialist societies are bigger with a widening of the lower tail. This is taken as evidence of the unhappiness of those who lost out in the transition. He finds the same in China. These authors are explicitly using the inequality of SWB to pick up the broader texture of inclusion at a given levels of mean income and growth.

Indeed, similar support is introduced at the end of the paper. Having affirmed a compression of SWB with income growth that is robust across levels of development, the authors conclude:

> This pattern is contradicted for the two countries in the dataset with the greatest economic growth rates: China and Korea. Perhaps, exceptional economic growth rates do not lead to decreasing SWB inequality over time, as such growth rates might cause large changes that affect citizens' SWB in disparate ways.

This is exasperating but effective as an advertorial for this research program; a scene from a forthcoming attraction. In the main my comments have been directed at supporting the potential of the broader framework and arguing that there is more in the current attraction to make this case.

References

Easterlin, R.A. (2012) "Life Satisfaction of Rich and Poor Under Socialism and Capitalism," *International Journal of Happiness and Development*, vol. 1, pp. 112–26.

Veenhoven, R. (2005) "Return of Inequality in Modern Society? Test by Dispersion of Life-Satisfaction Across Time and Nations," *Journal of Happiness Studies*, vol. 6, pp. 457–87.

Index